









genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wag1, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Machitani, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multichipillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.

## FEATURES

## source

1.438  
location/Qualifiers

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="G930013K04"

/clone.lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B) cDNA"

/tissue\_type="mammary gland"

BASE COUNT 138 a 96 c 108 g 96 t

## ORIGIN

## alignment\_scores:

Quality: 193.00 Length: 42

Ratio: 4.707 Gaps: 1

Percent Similarity: 97.619 Percent Identity: 95.238

## alignment\_block:

US-08-973-363-6 x BB830730 ..

Align seg 1/1 to: BB830730 from: 1 to: 438

```
1 1leleuProaspProaspPlylsyProglAlalysGlnleuGlnth 17
|||||
300 ATTCTTCACATGATGCTGATTAACCAACCAACCAACAGCTTACGAC 349
17 fArgAlaSPtyrLeuIleuLeuSerArgaspLeuAlaLysArg 33
|||||
350 CCGTGCAGACTACCTCTCAACTAGCTAGCAGAGATCTGCACAAAAGA 399
34 GluAlaGlnArgLeuGlyGlyAla 41
|||||
400 GAGGCTCAGAGACTTGTGTGCG 423
```

seq\_name: gb\_est1:AL601246

## seq\_documentation\_block:

LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001

DEFINITION DKFZP313J1040.F1 313 (synonym: hlcc2) Homo sapiens cDNA clone

DKFZP313J1040.5', mRNA sequence.

ACCESSION AL601246

VERSION AL601246.1 GI:15164752

KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 430)

Blöcker, H., Boecker, M., Brandt, P., Mewes, W., Well, B. and Wiemann

S., Est (Blöcker, H., Boecker, M., Brandt, P., Mewes, W., Well, B. and

Wiemann, S.)

JOURNAL Unpublished (1999)

COMMENT Contact: Blöcker H

MIRS

Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email: [s.wiemann@dkfz-heidelberg.de](mailto:s.wiemann@dkfz-heidelberg.de);  
sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No si sequence available.  
This clone (DKFZP313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: [clone@rzpd.de](mailto:clone@rzpd.de).

## FEATURES

## source

1.430

location/Qualifiers

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKFZP313J1040"

/clone.lib="313 (synonym: hlcc2)"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Vector: pTriplex2; site\_1: SfiI; site\_2: SfiIB;  
cDNA-collection"

BASE COUNT 163 a 81 c 86 g 100 t

## ORIGIN

## alignment\_scores:

Quality: 174.00 Length: 41

Ratio: 4.579 Gaps: 1

Percent Similarity: 92.683 Percent Identity: 90.244

## alignment\_block:

US-08-973-363-6 x AL601246 ..

Align seg 1/1 to: AL601246 from: 1 to: 430

```
1 1leleuProaspProaspPlylsyProglAlalysGlnleuGlnth 17
|||||
218 ATTCTTCACATGATGCTGATTAACCAACCAACCAACAGCTTACGAC 267
17 fArgAlaSPtyrLeuIleuLeuSerArgaspLeuAlaLysArg 34
|||||
268 CCGTGCAGACTACCTCAATTAATTAATTAATTAATTAATTAATTAAT 317
34 LuAlaGlnArgLeuGlyGlyAla 41
|||||
318 AAGCT.....CTTTCTGTGCG 334
```

seq\_name: gb\_est1:AI890775

## seq\_documentation\_block:

LOCUS AI890775 547 bp mRNA linear EST 07-MAR-2000

DEFINITION wms95f11.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:2443725 3'

similar to SW:CH01\_HUMAN 014646 CHROMODOMAIN-HELICASE-DNA-BINDING

PROTEIN 1; mRNA sequence.

ACCESSION AI890775

VERSION AI890775.1 GI:5595939

KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 547)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/brp/image/image.html  
 Insert length: 1924 Std Error: 0.00  
 Seq primer: 400P from Glbco  
 High quality sequence stop: 418.  
 Location/Qualifiers

## FEATURES

source

1..347  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="NC1CGAP\_012"  
 /tissue\_type="moderately-differentiated endometrial  
 adenocarcinoma, 3 pooled tumors"  
 /lab\_host="DH10B"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: Salt;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.85 kb. Life Technologies catalog #:  
 11539-012"  
 BASE COUNT 114 a 118 c 85 g 230 t  
 ORIGIN

## alignment\_scores:

Quality: 174.00 Length: 41  
 Ratio: 4.579 Gaps: 1  
 Percent Similarity: 92.683 Percent Identity: 90.244

## alignment\_block:

US-08-973-363-6 x A1890775/rev ..

Align seg 1/1 to reverse of: A1890775 from: 1 to: 547

1 HleuProApaspproaplysypproglnalalyglnleuglnth 17  
 |||||  
 378 ATTCTTCGAGATGATCCCGATTAACCAACACACACACACTTCGACAC 329  
 17 TATGALAApTYrleuIlleuSerATgAspIleuAluysarg 34  
 |||||  
 328 CCGTGGAGACTACCTCATCAATTACTTAGAGAGTCTGCAGAAAAAG 279  
 34 TAAAGlnAArgLeuCyGlyAla 41  
 |||||  
 278 AAGCT.....CTTCTGCTGCG 262

seq\_name: gb\_est1:AUI25712

## seq\_documentation\_block:

LOCUS AUI25712 866 bp mRNA linear EST 23-OCT-2000  
 DEFINITION AUI25712 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
 sequence.

ACCESSION AUI25712 GI:10950428

## KEYWORDS

SOURCE

ORGANISM

## REFERENCE

AUTHORS

## TITLE

## JOURNAL

## COMMENT

Human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 866)  
 Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
 Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
 Isogai,T.  
 HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
 Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,  
 Y., Sugano,S., Isogai,T.)  
 Unpublished (2000)  
 Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yama, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3951  
 Fax: 81-438-52-3952  
 Email: genomics@hri.co.jp  
 HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix

Research Institute; cDNA library construction; Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 Location/Qualifiers

## FEATURES

source

1..866  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="NT2RM4002061"  
 /clone\_id="NT2RM4"  
 /cell\_type="teratocarcinoma"  
 /cell\_line="NT2"  
 /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
 precursor cells"  
 BASE COUNT 312 a 149 c 196 g 207 t 2 others  
 ORIGIN

## alignment\_scores:

Quality: 174.00 Length: 41  
 Ratio: 4.579 Gaps: 1  
 Percent Similarity: 92.683 Percent Identity: 90.244

## alignment\_block:

US-08-973-363-6 x AUI25712 ..

Align seg 1/1 to: AUI25712 from: 1 to: 866

1 HleuProApaspproaplysypproglnalalyglnleuglnth 17  
 |||||  
 450 ATTCTTCGAGATGATCCCGATTAACCAACACACACACACTTCGACAC 499  
 17 TATGALAApTYrleuIlleuSerATgAspIleuAluysarg 34  
 |||||  
 500 CCGTGGAGACTACCTCATCAATTACTTAGAGAGTCTGCAGAAAAAG 549  
 34 TAAAGlnAArgLeuCyGlyAla 41  
 |||||  
 550 AAGCT.....CTTCTGCTGCG 566

seq\_name: gb\_est2:BE895133

## seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
 DEFINITION 60143606F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
 mRNA sequence.

ACCESSION BE895133 GI:10358221

## KEYWORDS

SOURCE

ORGANISM

## REFERENCE

AUTHORS

## TITLE

## JOURNAL

## COMMENT

Human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 1028)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: ceasbs-remail.nih.gov  
 Tissue Procurement: ATCC/CDTD/DTF  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Invitrogen, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: ILAM9753 row: h column: 16  
 High quality sequence stop: 488.  
 Location/Qualifiers

## FEATURES

source

1..1028  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="IMAGE:3921087"  
 /clone\_id="NIH\_MGC\_72"

```

/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: Skin; Vector: pCMV-SportC; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."

```

BASE COUNT 387 a 205 c 238 g 198 t

ORIGIN

alignment\_scores:

	Quality:	174.00	Length:	41
	Ratio:	4.579	Gaps:	1
Percent Similarity:	92.683	Percent Identity:	90.244	

alignment\_block:

US-08-973-363-6 x BE895133 ..

Align seg 1/1 to: BE895133 from: 1 to: 1028

```

1 TleupProaspaspproaspplyslysglproglinalalysglneuglnth 17
|||||
131 ATTCCTTCACATGATCCCGATTAACCAACCAACCAACGTTGCAGAC 180
|||||
17 ratgalaasptyrleuilelylseuleuseararaspdeulalysarg 34
|||||
181 CCGTGCAGACTACCTCATCAATTAATTAGTAGAGATCTGCACAAAAAAG 230
|||||
34 lualaglnargleucysglyala 41
|||||
231 AACCT.....CTTTCGCTGCG 247

```

seq\_name: gb\_est1:AL659353

seq\_documentation\_block:

LOCUS	AL659353	593 bp	mRNA	linear	EST 13-DEC-2000
DEFINITION	AL659353 XGC-neurula silurana tropicalis cDNA clone TNeu045e20 5' ,				
ACCESSION	AL659353				
VERSION	AL659353.1	GI:1672995			
KEYWORDS	EST.				
SOURCE	Western clawed frog.				
ORGANISM	Silurana tropicalis				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Silurana.				
AUTHORS	1 (bases 1 to 593)				
TITLE	Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.				
JOURNAL	Sanger Xenopus tropicalis EST project 2001 (10-2001)				
COMMENT	Unpublished (2001) Contact: Huckle E Sanger Centre Hinxton, Cambridgeshire, CB10 1SA, UK Email: tropesanger.ac.uk Sanger Xenopus tropicalis EST project 2001 TROPICALIS_SEQUENCE_ID: TNeu045e20.sp6 Sequencing primer: SP6 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn. Location/Qualifiers 1..593 /organism="Silurana tropicalis" /db_xref="taxon:8364" /clone="TNeu045e20" /clone_lib="XGC-neurula" /dev_stage="neurula" /lab_host="Escherichia coli DH10B" /note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; CDNA was oligo dT primed from 5ug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."				

FEATURES

Source

BASE COUNT 119 a 134 c 114 g 226 t

ORIGIN

```

alignment_scores:
    Quality: 172.00          Length: 38
    Ratio: 4.649            Gaps: 0
Percent Similarity: 97.368   Percent Identity: 86.842

alignment_block:
05-08-973-363-6 x AL659353/rev ..

Align seg 1/1 to reverse of: AL659353 from: 1 to: 593

      1  TleleProaspasproaspyslystysProglnAlalalyGlnlenglnh 17
        |||||||
      346 ATTATTACAGATGATCCAGATAAGAAGCCCCCAAGCCAAACACTTACAGAC 297
        |||||||
      17  rArgAlAsPtyrLeuIlleLysleuIeuSerArgrgaspleuAlalAlysarg 34
        |||||||
      296 CAGAGCTGCATCTCATTAACTTTTCATTAAGAATCTTGTTAGGAAG 247
        |||||||
      34  lualaglnargleu 38
        |||||||
      246 AAGCACACAAGACTT 233

seq_name: gb_est1:AL644594

seq_documentation_block:
LOCUS       AL644594                645 bp    mRNA    linear    EST 07-NOV-2001
DEFINITION  AL644594 XGC-egg Silurana tropicalis cDNA clone l1e1d12 5', mRNA
sequence.
ACCESSION   AL6444594
VERSION     AL644594.1 GI:16796719
KEYWORDS    EST.
SOURCE      western clawed frog.
ORGANISM    Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 645)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
JOURNAL
COMMENT     Contact: Huckie E
              Sanger Centre
              Hinxton, Cambridgeshire, CB10 1SA, UK
              Email: trop@sanger.ac.uk
              Sanger Xenopus tropicalis EST project 2001
TROPHICALIS_SEQUENCE_ID: l1e1d12.pic
Sequencing primer: PLC
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
    1..645
     /organism="Silurana tropicalis"
     /db_xref="taxon:8364"
     /clone="l1e1d12"
     /clone_lib="XGC-egg"
     /dev_stage="egg"
     /lab_host="Escherichia coli XL1-blue"
     /note="Vector: pCS107; Site.1: EcoRI; Site.2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"
BASE COUNT      222 a      125 c      156 g      141 t      1 others
ORIGIN

```

alignment\_block:

```
alignment_block:
```

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

```

2 LeuProAspProAspLysLysProGlnAlaLysGlnLeuGlnTrp 18
|||||
674 CTGCGGATATCTCTGATAGAGCCCTCAGCCCAACAGTTACGACAG 625
18 GAlaAspTyrLeuLeuLysLeuSerArgAspLeuAlaLysArgGln 35
|||||
624 AGCCDCTACTCATCTGCTGAGCAAGACCTGCGCAGAAAGAG 575
35 laGlnArgLeuCysGly 40
|||||
574 CCCACAGACGAGSGGGC 558

```

seq\_name: gb\_est1:AW996787

seq\_documentation\_block:

LOCUS AW996787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW996787

VERSION AW996787.1 GI:8257021

KEYWORDS

SOURCE human.

ORGANISM human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 337)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,  
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,  
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare  
, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
Simpson, A. J.

REFERENCE

AUTHORS

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL  
MEDLINE  
COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-230

200-102-d03&tl=2000-02-23&tl=1)

Seq primer: puc 18 forward

High quality sequence start: 2

High quality sequence stop: 337.

Location/Qualifiers

1..337

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BN0047"

/dev\_stage="Adult"

/note="Organ: breast.normal; Vector: puc18; Site.1: SmaI;  
Site.2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT

83 a 59 c 72 g 123 t

ORIGIN

alignment\_scores:

Quality: 139.50

Ratio: 4.103

Percent Similarity: 94.444

Length: 36

Gaps: 2

Percent Identity: 88.889

alignment\_block:

US-08-973-363-6 x AW996787/rev ..

Align seg 1/1 to reverse of: AW996787 from: 1 to: 337

```

1 lLeuProAspProAspLysLysProGlnAlaLysGlnLeuGlnTrp 17
|||||
105 ATTCTCCAGATGATCCCGATAAAAAACCAACCAACAGCTTGACAG 56
17 rArgAlaAspTyrLeuLeuLysLeuSerArgAspLeuAlaLysArg 33
|||||
34 GlnAla 35
|||||
8 GAAGCT 3

```

seq\_name: gb\_est1:AW997058

seq\_documentation\_block:

LOCUS AW997058 666 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW997058

VERSION AW997058.1 GI:8257292

KEYWORDS

SOURCE human.

ORGANISM human sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 666)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,  
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,  
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare  
, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
Simpson, A. J.

REFERENCE

AUTHORS

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL  
MEDLINE  
COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-150

400-152-c03&tl=2000-04-15&tl=1)

Seq primer: puc 18 forward

High quality sequence start: 19

High quality sequence stop: 678.

Location/Qualifiers

1..686

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BN0047"

/dev\_stage="Adult"

/note="Organ: breast.normal; Vector: puc18; Site.1: SmaI;  
Site.2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT

165 a 154 c 126 g 241 t

ORIGIN

## alignment\_scores:

Quality: 139.00 Length: 35  
Ratio: 3.971 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 94.286

## alignment\_block:

US-08-973-363-6 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

```

1 11leuPProkapsProkapsLysLysProGlnAlaLysGlnLeuGlnTh 17
|||||
127 ATTCTTCAGATGATCCGATTA.AAACCAAGCAAAACGTTGCAGAC 79
|||||
17 rArgAlaAspTyrLeuLeuLysLeuSerArgAspLeuAlaLysArg 34
|||||
78 CCGTCAGACTACCTCATCAATTAATTACTAGAGACTCTGCAAAAGAA 29
|||||
34 LuAla 35
|||||
28 GCTCT 24

```

seq\_name: gb\_gss:CNS05T9J

seq\_documentation\_block:

LOCUS CNS05T9J 1122 bp DNA linear GSS 26-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
042M09 of library C from Tetraodon nigroviridis, genomic survey  
sequence.  
ACCESSION AL352864.1 GI:8246657  
VERSION AL352864  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 1122)  
Roest-Crolius H., Jallion O., Dasilva C., Fizames C., Fisher C.,  
Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and  
Weissenbach J

TITLE Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 1122)  
Roest-Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,  
Bernot A., Fizames C., Wincker P., Brothier P., Quetier F.,  
Saurin W. and Weissenbach J.

TITLE Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
JOURNAL Unpublished  
AUTHORS 3 (bases 1 to 1122)  
Genoscope.

JOURNAL Direct Submission  
AUTHORS Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
COMMENT This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
http://www.genoscope.cns.fr/Tetraodon.  
FEATURES  
source  
1..1122  
Location/Qualifiers  
/organism="Tetraodon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_1lb="C"  
/clone\_1lb="C"  
/note="Genoscope sequence ID : C08C042AG05C1-end : T7"

BASE COUNT 303 a 260 c 301 g 246 t 12 others  
ORIGIN

alignment\_scores: Quality: 121.00 Length: 29

Ratio: 4.172 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 75.862

## alignment\_block:

US-08-973-363-6 x CNS05T9J ..

Align seg 1/1 to: CNS05T9J from: 1 to: 1122

```

1 11leuPProkapsProkapsLysLysProGlnAlaLysGlnLeuGlnTh 17
|||||
835 ATTCTTCAGATGATCCGATTA.AAACCAAGCAAAACGTTGCAGAC 884
|||||
17 rArgAlaAspTyrLeuLeuLysLeuSerArgAsp 29
|||||
885 CAGAGCTGAGTATCTCTCAAGCTGCTGAAAGAGAC 921
|||||

```



...



```
source 1. 6608
        /organism="unidentified"
        /db_xref="taxon:32644"
BASE COUNT 2289 a 1207 c 1459 g 1653 t
ORIGIN

alignment_scores:
    Quality: 205.00      Length: 41
    Ratio: 5.000         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x A58691 ..

Align seg 1/1 to: A58691 from: 1 to: 6608

1 1lleuProaspasProaspLysProGlnAlaLysGlnLeuGln 17
|||||
4080 ATTTTACCTGATGATCCAGACAAAGAACCCAGCAAGCCTTGACAGAC 4129
17 fArgAlaAspTyrLeuLysLeuAsnLysAspLeuAlaArgLysG 34
|||||
4130 CCGTGCAGACTACTCTATTAAATTACTGAAATAAGACCTTGCAAGAAG 4179
34 luAlaGlnArgLeuAlaGlyAla 41
|||||
4180 AAGCACAAGAGCTTGCTGTGCA 4202

seq_name: gb_ov:AF004397

seq_documentation_block:
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997
DEFINITION Gallus chromo-helicase-DNA-binding on the 2 chromosome
protein, variant with hydrophilic domain, (CHD-2) mRNA, complete
cds
ACCESSION AF004397
VERSION AF004397.1 GI:2501845
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
1 (bases 1 to 6872)
REFERENCE
AUTHORS Griffiths,R. and Korn,R.M.
TITLE A CHD1 gene is 2 chromosome linked in the chicken Gallus domesticus
JOURNAL Gene 197 (1-2), 225-229 (1997)
MEDLINE 97473516
REFERENCE
AUTHORS Griffiths,R. and Korn,R.M.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,
Glasgow G12 8QQ, UK

FEATURES
source
1. 6872
/organism="Gallus gallus"
/db_xref="taxon:9031"
1. 6872
/gene="CHD-2"
228.5654
/gene="CHD-2"
/function="role in chromatin architecture"
/note="CHD protein with hydrophilic domain"
/codon_start=1
/product="Chromo-helicase-DNA-binding on the 2 chromosome
protein"
/protein_id="AAC60282.1"
/db_xref="GI:2501846"
/translation="MNGHSDSESVNNSGESSRSDDSGASGSGSSGSSGSSGSS
SQSGSDSESSGSGSSESDTRREKKQVAKPPKADGSEFMKSPSITLAVQSAV
LKQOQOQKAAASDSDGSESSSDSDSSSEFKKKHKHDKEDQMGSGSGSVGTGS
DSESEADGDKSSCESESDYEPKKNVKSRRPSRIKPSGKKSTGCKRQDLSSEEE
```

```
DDDEDYDKRGRQATVNVYSKEAEETKTDSDDLVEGGEVDPQTEDEFEETIEKFM
SRIGRGATGASTTYYAAVEDGDPNAGFEKSELEGIQYLILKMGKSHIHNTWEET
LKQONVKGKNDKNDKNDKNDKNDKNDKNDKNDKNDKNDKNDKNDKNDKNDKNDK
AHSNOKSAGAPDYCKKNOGIPYSCSDGALITAKKQARIDEPYRNOSKTPFKD
KYLKQRPYVALKKQPSYIGHESELEBDYQNLGNLHSHWCGNSCLIDENGLG
KTQITSPNLVLEHEDLYGPFLRPLSTLTSWQREIQVAPQNAVYILKLVYGS
MIRHEMMHPQTRKLFNILLTYEILILDKSFLGLMNAFIGVDEARLKNDLSLY
RTLIDPKSNHRLITGTPLQNSLKEIWSLHIFIMPEKFSWDEFEHKGREYAS
LHKELEPFLRRVKKADVERSLAPKAVEOLIRMSALOKOYWMILTRNYKALSKSG
STSGFLNIMELKCCNCYLILPPDDNPEFYKQALOHILRSSGKLLDLKLRLR
EKGNYLIFSQYVRMLDILAEYLKTRQPPQDLDSIKGELRKAOLDHFNAGSDEFC
FLSTRAGGLGINLASADTVLEDSDMWPNQDAQARHRIQKQVNYIKLVYGS
VEEDILERRAKKMVLDHLYIQMDITGKTVLITGTPSSSTPNNKEISAILKFAE
LEKPEGEEOPEQMDIDELIKRAETRENEPGLVGEDLLSQFYVANSNDEDDIE
LEPERNSRWMEELIPESORRIEERERKELEIYMLPRMRCAKOISPNSEGRSR
SRYSQSDSTTEKRPKRRPRRTIRENTKGSQDAIRFRTISYKPKPLERLD
AVARDELVDKSETDLRKGELVHNGCTIKALDNSGQBRAGRLGKKGPFRTISG
OVNARLVISHEELAPLRSKIPSDPEERKRYIPQHTAAHEDIDWGEDSNLLVG
YEGYGSWEMIKMDPDLSTOKILPDDPKQQAQLOTRADYLYLILNKDLARBAQ
FLAGAGNSKRRRTNRKNNKMKASKIKEELSDSPSEKSEDEDEEDNDKEIYSVK
HLAKKITEKENEKEPEPDIGIKKEAEERETKEKREBELKREKEDKLEKD
NKEKRENNKESTOKEKYEKVEKSENERKSKTPLDTPYHITATSPRPLSE
BSEELHQKTFSTCKERMRPVKALQOLDREPKGSEBQLETRQCLIKIGHTTECL
KEYTPEQIKQWRKNMLFVSKFTEFDARKLKLKYLKHAIKRQESQOHNDQNTSSNV
THVINRPDVERLKEETNHDSSRDSYSDRHLISOYHDKHDKHODAYKKSDDRPPY
SAFNSKDHDRMDHYKODSRYSDSKHRLDHRSDRSNLEGMILKDSRSDHRS
SDRHSKDHRSSTSEYSHKSSRDYRHSQWQMDHRASSGSPRPLDQSPYGRSPLG
HSPREHSDHKSTPEHTWSSKRT"
4341..14604
/misc-feature
/gene="CHD-2"
/note="short insert found in longer variant mRNA of CHD-2"
BASE COUNT 2446 a 1223 c 1520 g 1683 t
ORIGIN

alignment_scores:
    quality: 205.00      Length: 41
    Ratio: 5.000         Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x AF004397 ..

Align seg 1/1 to: AF004397 from: 1 to: 6872

1 1lleuProaspasProaspLysProGlnAlaLysGlnLeuGln 17
|||||
4080 ATTTTACCTGATGATCCAGACAAAGAACCCAGCAAGCAGCTACAGAC 4129
17 fArgAlaAspTyrLeuLysLeuAsnLysAspLeuAlaArgLysG 34
|||||
4130 CCGTGCAGACTACTCTATTAAATTACTGAAATAAGACCTTGCAAGAAG 4179
34 luAlaGlnArgLeuAlaGlyAla 41
|||||
4180 AAGCACAAGAGCTTGCTGTGCA 4202

seq_name: gb_pat:A58684

seq_documentation_block:
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 3 from patent WO9639505.
ACCESSION A58684
VERSION A58684.1 GI:3714247
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 153)
AUTHORS Griffiths,R. and Tiwari,B.
TITLE AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN
BIRDS
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;
```

```

COMMENT      ISIS INNOVATION (GB)
FEATURES     other publication AU 5906996 961224.
source       1..153
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT   58 a      40 c      31 g      24 t
ORIGIN
Alignment_scores:
Quality:     190.00      Length:    51
Ratio:       4.634       Gaps:      1
Percent Similarity: 80.392 Percent Identity: 80.392

seq_name: gb_pat:A58685
seq_documentation_block:
LOCUS         A58685              153 bp      DNA          linear      PAT 06-MAR-1998
DEFINITION   Sequence 4 from Patent WO9639505.
ACCESSION    A58685
VERSION      A58685.1 GI:3714248
KEYWORDS
SOURCE       unidentified.
ORGANISM     unidentified.
REFERENCE    1 (bases 1 to 153)
AUTHORS     Griffiths,R. and Tiwari,B.
TITLES      AVIAN GHG GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL      Patent: WO 9639505-A 4 12-DEC-1996;
COMMENT      ISIS INNOVATION (GB)
FEATURES     Other publication AU 5906996 961224.
source       1..153
            /organism="unidentified"
            /db_xref="taxon:32644"
BASE COUNT   56 a      36 c      31 g      30 t
ORIGIN
Alignment_scores:
Quality:     190.00      Length:    51
Ratio:       4.634       Gaps:      1
Percent Similarity: 80.392 Percent Identity: 80.392

seq_name: gb_ro:MUSCHDIX
seq_documentation_block:
LOCUS         MUSCHDIX             5349 bp      mRNA          linear      ROD 19-SEP-1996
DEFINITION   Mouse DNA-binding protein (CHD-1) mRNA, complete cds.
ACCESSION    L10410 X66028
VERSION      L10410.1 GI:455014
KEYWORDS     DNA binding protein.
SOURCE       house mouse.
ORGANISM     Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    Delmas,V., Stokes,D.G. and Perry,R.P.
AUTHORS      1 (bases 1 to 5349)
TITLE        A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)
MEDLINE      93211972
REFERENCE    2 (bases 1 to 5349)
AUTHORS      Perry,R.P.
TITLES      Direct Submission
JOURNAL      Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA
COMMENT      On Feb 16, 1994 this sequence version replaced g1:293322.
FEATURES     Location/Qualifiers
source       1..5349
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /cell_type="plasmacytoma"
            .5349
gene         1..5349
CDS          171..5306
            /gene="CHD-1"
            /product="DNA-binding protein"
            /protein_id="AAB08486.1"
            /db_xref="GI:455015"
            /translation="MNGSHDESVRNKSGSSQSDGCGASGSAGSSGSSGSSDGSS
            SGGSSDLSGSGDSQSSESDPTREKNVOAKPPVDGAERKMSPSLTAVERAMLM
            RKPOQAQQQRASNSNGSEEDSSSEDSSSGAKRKHNDDMDMGSGSPQLGASQ
            SDSEEFEEDRSKCDGTESDYEPKRNKRVSRRPNRSKNCK ILGKKRKQSDSEDE
            DDBDYNDKRSRRQATVANYSYKEDBEKMKDSDLDLEGVKGQWQPDEDFET
            DCVGVRGATGCTTTIYAVADGDNAEFENKEPDD IGYCIKKGSHSIHMTVEETE
            TLQOQVRGMKLIDNYKKKOQCFEKRWLNASPDVVYVNCQGLDTDLKQOYIER
            TASHNKSAGCLIPYCYKMGGLPYSCEWDGALISKPCICIDYESNRQSTPEPK
            DCVLKORPRVALKKOPSTIGSHGLELBYOLNIMLAHSMCKNSCIILADMGK
            KTIOTIFSLTYLFERHOAYGPILLVPLSTLSMRQELORMASOMNAVYIGDIINS
            NMTRHEMMHQOTRYKENTLLITYETLLDKAFGLMAPIQVDPAHRLKNDISLL
            YKTLIDPKSNRIILITGTPLONSIKELISLIHRIMPRESSWDPFEFGKRGVCYA
            SLKLELPLELRVKRYKVESLPKAVVDILRMKSLKQYTKITLRVYKLSKGSF
            GSTSGFLIMMEYKCCNCTILRPDNNEFTNKQDALHLKSSGKVLIDLKLRIRI
            REGRNVLTIESQWRMLDILAEIKTROFPQRLDSTIGELRKQALDHFNBSGEDE
            CFLTSTRAGGAG INVASLOTIVYTFSDNNPONDIAQAFAIRTOOKOVNIITLVYG
            SVBEDLTERRAKKMWLDHLYDMOTTCGLVTGSAESSSTPNKBELSAITLKGAB
            ELFREPEGEDEPQMDIELIKRAETHENDPJSVDELISQFVAVNSWMDDDIT
            ELPERNSKNMBEIIPEQORKLEEBEOKELBEIYMPPRRNCAQISFNSBGRSR
            RSRVYSGSDSDSISERRPKKRGRRPTIPBNIKGFSDAETIRINSYKKGCPGERI
            DATAEALPYVSKSETDLRRIGELVINGCVALKDSSSGTERAGRGKVGKPTFISG

```





```

seq_name: gb_pr:AC012624
26832 AAGCT.....CTTCTGGTGCG 26816

34 lu1aGlnArgrleuaAglYAla 41
||||| |||||
||||| .....

17 rrrgaLaasPTyrLeuIleLysLeuLeuSnlYsaAspleuAlaArGLysG 34
||||| .....
CCGTGCAGACTCACCTCATCAATTAATTACTTAGTAGAGATTCTTGCAAAAAAG 26833

seq_documentation_block:
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:14993679
KEYWORDS HPG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
source
location/Qualifiers
1..134365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN
alignment_scores:
Quality: 170.00 Length: 41
Ratio: 4.359 Gaps: 1
Percent Similarity: 95.122 Percent Identity: 85.366

Alignment Block:
US-08-973-363-7 x AC012624 ..

Align seg 1/1 to: AC012624 from: 1 to: 134365

1 ITeLeuPrOaSPaSPProCaSPyLSlYsProGlInAlaLySGInLeuGlnPh 17
||||| .....
118247 ATTCTTCACGANGATCCGATATAAAACCACAACGAAGTTGCCAGAC 118296

17 rrrgalaasPTyrLeuIleLysLeuLeuSnlYsaAspleuAlaArGLysG 34
||||| .....
118297 CCGTGCAGACTCACCTCATCAATTAATTACTTAGTAGAGATTCTTGCAAAAAAG 118346

34 lu1aGlnArgrleuaAglYAla 41
||||| .....
118347 AAGCT.....CTTCTGGTGCG 118363

seq_name: gb_htg:AC021449
```

```

seq_documentation_block:
LOCUS      AC021449             143079 bp    DNA    linear    HTG-10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION  AC021449
VERSION    AC021449.3  GI:10047806
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 143079)
            Barren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL    Unpublished
TITLE      2 (bases 1 to 143079)
AUTHORS     Barren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 143079)
            Barren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckery,R., Beda,F.,
            Bodusivsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepeel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            Deatrelano,P., Dewar,K., Domino,M., Doyle,M., Fennestor,J.,
            Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karalas,A., Klein,J.,
            Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,
            MacDonald,P., Margulis,N., McEwan,P., McGurt,A., McKernan,K.,
            McPheters,R., Meldrum,J., Menues,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
            Pierre,C., Pisani,C., Pollard,V., Raymond,C., Riley,R., Rotman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomas,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
JOURNAL     Direct Submission
COMMENT     Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02144, USA
            On Sep 10, 2000 this sequence version replaced gi:7407963.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIRB
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L5154
            Center clone name: 58_M_12
            ----- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 134743 bases at least Q40
            Consensus quality: 139227 bases at least Q30
            Consensus quality: 140814 bases at least Q20
            Insert size: 14400; agarose-fp
            Insert size: 142179; sum-of-contigs
            Quality coverage: 4.6 in Q20 bases; agarose-fp
            Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 38820: contig of 38820 bp in length
            * 38821 38920: gap of 100 bp
            * 38921 40411: contig of 1491 bp in length
            * 40412 40511: gap of 100 bp

```





Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366  
alignment\_block:  
US-08-973-363-7 x AC008531 ..  
Align seg 1/1 to: AC008531 from: 1 to: 145659  
1 11leuProaspProaspLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
46770 ATTCTGCACATATATCCCGATMAAAACACAGCAAAACAGTTGCAGAC 46819  
17 rAtgAlaAspTyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
46820 CCGTGAGACTACTCATCAATTAAGTAGAGAGATCTGCMAAAAMAG 46869  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
46870 AACCT.....CTTCTGTGTCG 46886  
seq\_name: gb\_hcg:AC091946  
seq\_documentation\_block:  
LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
PROGRES \*\*\*, 33 unordered pieces.  
ACCESSION AC091946  
VERSION AC091946.1 GI:14333882  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 544799  
Center clone name: RPCT-11\_36012  
-----  
Summary Statistics  
Consensus quality: 157767 bases at least Q40  
Consensus quality: 180259 bases at least Q30  
Consensus quality: 184175 bases at least Q20  
Estimated insert size: 204590; agarose-IP estimation  
Quality coverage: 5.38 in Q20 bases; agarose-IP estimation  
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1358: contig of 1358 bp in length  
\* 1359 1458: gap of unknown length  
\* 1459 2667: contig of 1209 bp in length  
\* 2668 2767: gap of unknown length  
\* 2768 3823: contig of 1056 bp in length  
\* 3824 3923: gap of unknown length

\* 3924 5920: contig of 1997 bp in length  
\* 5921 6020: gap of unknown length  
\* 6021 7494: contig of 1474 bp in length  
\* 7495 7594: gap of unknown length  
\* 7595 9096: contig of 1502 bp in length  
\* 9097 9196: gap of unknown length  
\* 9197 10260: contig of 1064 bp in length  
\* 10261 10360: gap of unknown length  
\* 10361 12460: contig of 2100 bp in length  
\* 12461 12560: gap of unknown length  
\* 12561 14611: contig of 2051 bp in length  
\* 14612 14711: gap of unknown length  
\* 14712 16381: contig of 1670 bp in length  
\* 16382 16481: gap of unknown length  
\* 16482 17968: contig of 1487 bp in length  
\* 17969 18068: gap of unknown length  
\* 18069 20434: contig of 2366 bp in length  
\* 20435 20534: gap of unknown length  
\* 20535 23515: contig of 2981 bp in length  
\* 23516 23615: gap of unknown length  
\* 23616 27563: contig of 3948 bp in length  
\* 27564 30987: gap of unknown length  
\* 30988 31087: contig of 3324 bp in length  
\* 31088 36065: gap of unknown length  
\* 36066 36165: contig of 4978 bp in length  
\* 36166 40978: gap of unknown length  
\* 40979 41078: contig of 4813 bp in length  
\* 41079 45663: contig of 4585 bp in length  
\* 45664 45763: gap of unknown length  
\* 45764 51745: contig of 5982 bp in length  
\* 51746 51845: gap of unknown length  
\* 51846 57359: contig of 5514 bp in length  
\* 57360 57459: gap of unknown length  
\* 57460 67881: contig of 10422 bp in length  
\* 67882 74132: gap of unknown length  
\* 74133 74232: contig of 6151 bp in length  
\* 74233 79795: contig of 5563 bp in length  
\* 79796 79895: gap of unknown length  
\* 79896 87511: contig of 7616 bp in length  
\* 87512 92791: gap of unknown length  
\* 92792 92891: contig of 5180 bp in length  
\* 92892 102794: gap of unknown length  
\* 102795 102894: contig of 9903 bp in length  
\* 102895 110866: gap of unknown length  
\* 110867 110966: contig of 7972 bp in length  
\* 110967 117571: gap of unknown length  
\* 117572 117671: contig of 6665 bp in length  
\* 117672 123738: gap of unknown length  
\* 123739 123838: contig of 6067 bp in length  
\* 123839 130583: gap of unknown length  
\* 130584 130683: contig of 6745 bp in length  
\* 130684 141544: gap of unknown length  
\* 141545 141644: contig of 10861 bp in length  
\* 141645 169109: gap of unknown length  
\* 169110 169209: contig of 27465 bp in length  
\* 169210 193446: gap of unknown length  
\* 193446 24237: contig of 24237 bp in length.

## FEATURES

## Source

1.193446  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-36012"  
/clone\_lib="RPCT human BAC library 11"  
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others  
ORIGIN

## alignment\_scores:

Ratio: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1





OM of: US-08-973-363-7 to: N\_Geneseq\_032802: out\_format: pfs  
Date: Aug 3, 2002 4:34 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODEL=frame-p2n model -DPV=xlh  
-O=/cgn2.1/USPTO.spool/US08973363/runat.01082002.080050.18606/app-query.fasta.1.638  
-DB=N\_Geneseq\_032802 -OPMT=fastap -SUFFIX=p2n.rng -GAPOP=2.000  
-GAPEXT=4.000 -MINMATCH=0.100 -DOEXT=0.000 -DOEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELOP=7.000 -SHAP=1 -MATH=blsnum6  
-RANS=blsnum4 cdi LIST=45 -DOCLIG=200 -RHS\_SCORE=pcr  
-RHS\_MAX=100 -RHS\_MIN=0 -MODE=LOCAL -OVERRIDE=pts  
-NOR=ext HEADSIZE=500 -NLEN=0 -MAXLEN=200000000  
-USER=US08973363.ecgnl\_186 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_ALPXY -WAIT -THREADS=1

## Search information block:

Query: US-08-973-363-7  
Query length: 41  
Database: N\_Geneseq\_032802:  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 523.620000

## score\_list:

Sequence	Strid	Orig	ZScore	Escore	Len	Documentation
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	205.00	553.01	1.9e-22			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	205.00	537.19	1.4e-21			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	190.00	531.94	2.8e-21			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	190.00	531.94	2.8e-21			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	182.00	509.45	5.0e-20			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	172.00	481.33	1.8e-18			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	129.00	339.84	1.5e-10			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	122.00	330.84	4.4e-11			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	122.00	330.84	4.4e-11			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	88.00	208.82	0.0028			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	88.00	208.82	0.0028			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	81.00	204.26	0.0050			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	60.00	141.72	15.15			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	58.50	109.23	977.07			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	57.50	130.55	63.48			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	56.50	142.10	14.43			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	56.50	140.09	18.67			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	56.50	127.73	91.07			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	56.00	127.46	94.35			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	56.00	127.46	94.35			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	55.00	134.90	36.33			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	55.00	133.58	43.06			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	55.00	125.08	127.98			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.50	122.06	188.53			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.50	118.31	904.92			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.50	112.28	660.71			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.00	140.29	16.01			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.00	137.71	23.33			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.00	130.76	61.75			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.00	123.22	162.46			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.00	114.85	475.48			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.00	113.74	548.43			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.00	107.22	1.3e+0			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.00	103.61	2.0e+0			
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754	54.00	103.07	2.2e+03			

/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.AAF90033 - 54.00 96.59 4.9  
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.AAF90032 - 54.00 94.37 6.6  
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001A.DAT.AAF66431 + 54.00 73.75 9.1  
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.ABL26559 + 53.00 131.84 53.  
/SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA2001B.DAT.AAS74806 + 53.00 128.52 82.

## seq\_name: /SIDSI/gcgdata/hold-geneseq/geneq-emb1/NA1997.DAT.AAT42754

seq\_documentation block:

ID AAT42754 standard; CDNA: 1316 BP.

AC AAT42754:

DN 12-MAR-1997 (first entry)

DE Chicken CHD-W gene (partial sequence).

KW Bird; sex determination: chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

PN WO9639505-A1.

PD 12-DEC-1996.

PE 05-JUN-1996: 96WO-GB01341.

PR 06-JUN-1995: 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B:

WP1: 1997-043127/04.

PT Avian chromodomain-helicase-DNA binding genes determine sex in

PS birds - used for sex determin. and to control sex of progeny

PS Claim 1: Fig 8: 76pp: English.

CC The chicken CHD-W gene (AAT42754) acting alone or in conjunction with

CC the closely related CHD-1A gene (AAT42751) is suggested to initiate

CC female development in birds. The sequence of CHD-W was deduced

CC from 2 clones isolated from a 10-day chick embryo library using

CC a fragment of the CHD-1A gene as a probe. The CHD-W gene is

CC located on the W chromosome. Probes based on CHD-W and CHD-1A give

CC a W chromosome-specific signal on hybridisation to genomic DNA of a

CC nucleic acids can also be used to control the sex of progeny of a

CC bird.

XX Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other:

XX Alignment\_scores:

XX Quality: 205.00 Length: 41

XX Ratio: 5.000 Gaps: 0

XX Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-7 x AAT42754

Align seg 1/1 to: AAT42754 from: 1 to: 1316

1 TleleuProasPaspProaspLysPscGlnAlaLysGlnLeuGlnth 17

1138 ATTTTACCTGATGATTCAGATTAAGAAACCCAGGCTAACCAAGTACAGAC 1187

17 TATGAlaAPfyrleuLleuLysLeuLysLeuLysLeuLysLeuLysLeuLys 34

1188 TCGTGAAGATTACCTCATTAATTAATTAAGACCTTCAGAAAGAG 1237

34 luA1aGlnArgLeuAlaGlyAla 41  
 ||||||||||||||||||  
 1238 AAGCAGACAGACTGCTGTGTGCA 1260

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751

seq\_documentation\_block:

ID AA142751 standard; cDNA; 6608 BP.

AC AA142751;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-1A gene.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

Key Location/Qualifiers

FT CDS 228..5390

FT /\*tag= a

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

XX Claim 1; Fig 5; 76pp; English.

CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with

CC the closely related CHD-1A gene (AA142751) is suggested to initiate

CC female development in birds. The sequence of CHD-1A was deduced

CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA

CC library using a great tit CHD-W sequence (see also AA142755) as probe.

CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1

CC gene (see also AA142756-57). It is located on an autosome or Z

CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-

CC specific signal on hybridisation to genomic DNA of a non-ratite

CC bird and can be used for sex determin. of a bird. CHD-1A nucleic

CC acids can also be used to control the sex of the progeny of a bird.

XX Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores:

Quality: 205.00 Length: 41  
 Ratio: 5.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AA142751 ..

Align seg 1/1 to: AA142751 from: 1 to: 6608

1 ltleuProaspProaspLysProGlnAlaLysGlnLeuIntH 17

4080 AATTTTCCTGATGTCACAGACAAAGCCAGCAGACGCTACAGAC 4129

17 rArgAlaAspTyrLeuIleuLeuLeuAsnLysAspLeuAlaArgLysG 34

|||||  
 4130 CCGTGCAGACTACCTCATTAATTAAGTAAGACCTGCAGAAAGC 4179  
 34 luA1aGlnArgLeuAlaGlyAla 41  
 ||||||||||||||||||  
 4180 AAGCAGAAAGCTGCTGTGTGCA 4202

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142757

seq\_documentation\_block:

ID AA142757 standard; DNA; 153 BP.

AC AA142757;

DT 12-MAR-1997 (first entry)

DE Chick CHD-1A gene fragment.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KW CHD-1A; CHD-W; W chromosome; ss.

OS Gallus sp.

Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag= a

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

DR P-PSDB; AAW08147.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology

CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),

CC chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)

CC and the great tit CHD-W gene (see also AA142759). Translated amino

CC acid sequences of this region are provided in AAW08146-49. The

CC CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes

CC determine sex in birds and can be used to identify the sex of an

CC embryo, foetus etc. and to manipulate the sex of progeny.

XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment\_scores:

Quality: 190.00 Length: 51  
 Ratio: 4.634 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 80.392

alignment\_block:

US-08-973-363-7 x AA142757 ..

Align seg 1/1 to: AA142757 from: 1 to: 153

1 ltleuProaspProaspLysProGlnAlaLysGlnLeuIntH 17

4080 AATTTTCCTGATGTCACAGACAAAGCCAGCAGACGCTACAGAC 4129

17 rArgAlaAspTyrLeuIleuLeuLeuAsnLysAspLeuAlaArgLysG 34

```

|||||
1 ATTTACTGATGATTCACAGAACCCAGCAGACGACTACAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnTrpArgAlaAspTyrLeuLeu 24
51 CAAGAAACCCAGCAGCTACAGCTACAGCCCTGCAGACTACCTCACTTA 100
24 yslleuLeuAsnLysAspLeuAlaArgLysGlnAlaArgLeuAlaGly 40
101 AATTACTGAAATTAAGACCTTCAGAGAAAGAGACACAGACTGCTGGT 150
41 Ala 41
151 GCA 153

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758
seq_documentation_block:
ID AA142758 standard; DNA; 153 BP.
XX
AC AA142758;
XX
DT 12-MAR-1997 (first entry)
XX
DE Chick CHD-W gene fragment.
XX
KM Bird; sex determination: Chromodomain-Helicase-DNA binding 1;
XX CHD-1a; CHD-W; W chromosome; ss.
XX
OS Gallus sp.
XX
Key Location/Qualifiers
FH misc_difference 52..81 /tag= a
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PF 05-JUN-1996; 96WO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
DR WPI: 1997-043127/04.
XX
DR P-PSDB; AAM08148.
XX
PA Avian Chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
PS Claim 8; Fig 3; 76pp; English.
XX
CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAM08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,
XX foetus etc. and to manipulate the sex of progeny.
XX
SO Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment_scores:      Length: 51
Quality: 190.00      Gaps: 1
Ratio: 4.634

```

```

Percent Similarity: 80.392      Percent Identity: 80.392
alignment_block:
US-08-973-363-7 x AA142758
Align seg 1/1 to: AA142758 from: 1 to: 153

1 TlleuProAspAspProAsp.....7
1 ATTTACTGATGATTCACAGAACCCAGCAGCTACAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnTrpArgAlaAspTyrLeuLeu 24
51 CAAGAAACCCAGCAGCTACAGCTACAGCCCTGCAGACTACCTCACTTA 100
24 yslleuLeuAsnLysAspLeuAlaArgLysGlnAlaArgLeuAlaGly 40
101 AATTACTGAAATTAAGACCTTCAGAGAAAGAGACACAGACTGCTGGT 150
41 Ala 41
151 GCA 153

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759
seq_documentation_block:
ID AA142759 standard; DNA; 153 BP.
XX
AC AA142759;
XX
DT 12-MAR-1997 (first entry)
XX
DE Great tit CHD-W gene fragment.
XX
KM Bird; sex determination: Chromodomain-Helicase-DNA binding 1;
XX CHD-1a; CHD-W; W chromosome; ss.
XX
OS Parus major.
XX
Key Location/Qualifiers
FH misc_difference 52..81 /tag= a
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PF 05-JUN-1996; 96WO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
DR WPI: 1997-043127/04.
XX
DR P-PSDB; AAM08149.
XX
PA Avian Chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
PS Claim 8; Fig 3; 76pp; English.
XX
CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAM08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,

```



CC foetus etc. and to manipulate the sex of progeny.  
 XX Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

alignment\_scores:  
 Quality: 182.00 Length: 51  
 Ratio: 4.439 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 76.471

alignment\_block:  
 US-08-973-363-7 x AAT42759 ..

Align seg 1/1 to: AAT42759 from: 1 to: 153

```

1 1leuPProAspProasp..... 7
1 ATTTTACCTGATGATCCGATTAAGAAACACAGCAAGCAAGTTCAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
51 CAGAAACACAGCAAGCAAGTTCAGACCCGTCGACATTCATCTA 100
24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40
101 AATTAAGTAATTAAGACCTTGCAAGAAAGAAAGTGTCAAGACTTACTGT 150
41 Ala 41
151 GCA 153

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAT42756

seq\_documentation\_block:

ID AAT42756 standard; DNA; 153 BP.

XX AAT42756;

DT 12-MAR-1997 (first entry)

DE Mouse CHD-1 gene (bases 3855-977).

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KW CHD-1; CHD-W; W chromosome; ss.

OS Mus sp.

EH Key Location/Qualifiers

FT misc\_difference 52..81

FT /tag= a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 and are ignored in the translated amino  
 acid sequence given in Fig 3"

XX WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-CB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

DR P-PSDB; AAM08146.

XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
 CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene  
 CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
 CC Translated amino acid sequences of this region are provided in  
 CC AAM08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
 CC AAT42754-55) genes determine sex in birds and can be used to identify  
 CC the sex of an embryo, foetus etc. and to manipulate the sex of  
 CC progeny.

XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

alignment\_scores:  
 Quality: 172.00 Length: 51  
 Ratio: 4.195 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 70.588

alignment\_block:  
 US-08-973-363-7 x AAT42756 ..

Align seg 1/1 to: AAT42756 from: 1 to: 153

```

1 1leuPProAspProasp..... 7
1 ATTTTACCTGATGATCCGATTAAGAAACACAGCAAGCAAGTTCAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
51 CAAAAAACACAGCAAGCAAGTTCAGACCCGTCGACATTCATCTA 100
24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40
101 AACTACTTACGACGATCTTGCAAAAAAGAGAGCTCAGACACTTGTGCT 150
41 Ala 41
151 GCG 153

```

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV59280

seq\_documentation\_block:

ID AAV59280 standard; cDNA; 1311 BP.

XX AAV59280;

DT 14-DEC-1998 (first entry)

DE Altered telomere repeat binding factor 1 gene.

KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;  
 KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

OS Homo sapiens.

OS Synthetic.

EH Key Location/Qualifiers

FT CDS 1..1311

FT /tag= a  
 FT /product= "A-TRF"

XX WO9836066-A1.

PD 20-AUG-1998.

PF 13-FEB-1998; 98WO-US02765.

PR 04-FEB-1998; 98US-0018628.

PR 13-FEB-1997; 97US-0800264.  
 XX (UYNQ ) UNITV ROCKEFELLER.



PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0241617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249300
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250391
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251988
PR	05-DEC-2000	2000US-0256719
PR	06-DEC-2000	2000US-0251479
PR	08-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251868
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251989
PR	08-DEC-2000	2000US-0251990
PR	11-DEC-2000	2000US-0254097
PR	03-JAN-2001	2001US-0259678
PA	(HUMA-)	HDUMAN GENOME SCI INC.
PI	Rosen CA,	Barash SC, Ruben SW;
XX	WPI, 2001-502630/55.	
DR	P-PSDB; AAM93109.	
XX		
PT	Polynucleotides encoding digestive system antigens, useful for	
PT	diagnosing, treating, preventing and/or prognosing disorders of the	
PT	digestive system, particularly cancer and cancer metastases -	
XX	Claim 1; SEQ ID NO 1198; 966pp; English.	
XX	The present invention provides the protein and coding sequences of a	

CC	number of human digestive system antigens. These can be used in the
CC	diagnosis, treatment and prevention of digestive system disorders,
CC	including cancer, Meckel's diverticulum, bacterial or parasitic
CC	infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC	ulcerative colitis. The present sequence is a cDNA encoding a digestiv
CC	system antigen of the invention.
XX	
SD	Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;
	alignment_scores:
	Quality: 122.00 Length: 35
	Ratio: 4.067 Gaps: 0
	Percent Similarity: 85.714 Percent Identity: 71.429
	alignment_block:
	US-08-973-363-7 x AAK88882 ..
	Align seg 1/1 to: AAK88882 from: 1 to: 421
	1 TlelePProAspAspPProAspLysLysProGlnAlaLysGlnLeuGlnrh 17
	: : :           : : :
	110 ATTCGCGCGTGAGACAGATPAAAGCCTCAGGGGAAGCAGCTACAGAC 159
	17 rARgAlaAspTyrLleuLleLysLysLeuAsnLysAspLysAlaArgLysG 34
	: : :           : : :     : :
	160 CCGAGCGGATTACTGTGTGAAGCTGCTCAGAAAGCGTCTGAGAGAAGAGC 209
	34 lAA1a 35
	210 GGGCT 214
	seq_name: /SIDS1/gcgcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AA157603
	seq_documentation_block:
ID	AA157603 standard; cDNA: 421 BP.
XX	
AC	AA157603;
XX	
DT	19-OCT-2001 (first entry)
XX	
DE	Human colorectal cancer antigen cDNA SEQ ID NO: 67.
XX	
KW	Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss
XX	
OS	Homo sapiens.
XX	
PN	WO200155350-A1.
PD	
XX	02-AUG-2001.
PF	
XX	17-JAN-2001; 2001WO-US01350.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220363.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.



Percent Similarity: 85.714 Percent Identity: 71.429

alignment\_block:

US-08-973-363-7 x AAI57603 ..

Align seg 1/1 to: AAI57603 from: 1 to: 421

```
1 lileuProaspapProaspLysLysProGlnAlaLysGlnLeuGlnH 17
||||| :||| :||| :||| :||| :||| :||| :||| :|||
110 ATTCTGCGGTTGAGACAGATATAAGCTCAGGCGAGACGACGAC 159
17 rArgAlaAspTyrIleuIleLysLeuAsnLysAspLeuAlaArgLysG 34
||||| :||| :||| :||| :||| :||| :||| :||| :|||
160 CCGAGCGGATTACTTGTGAAGCTGTCAAGAAAGCTGTGAGAAAGG 209
34 luLa 35
|||
210 GGCGT 214
```

seq\_name: /SIDSI/gcgdata/hold-geneseq/genesegn-emb1/NA2001B.DAT:ABL06443

seq\_documentation\_block:

ID ABL06443 standard; CDNA: 6240 BP.

AC ABL06443;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.

KW Drosophila: developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB62340.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA

XX sequences (ABLI01840-ABLI6175) and the encoded proteins

XX (ABBS7737-ABBS72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

XX alignment\_scores:

Quality: 88.00 Length: 38

Ratio: 2.839 Gaps: 2

Percent Similarity: 81.579 Percent Identity: 55.263

alignment\_block:

US-08-973-363-7 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

```
1 lileuProaspapProaspLysLysProGlnAlaLysGlnLeuGlnH 17
||||| :||| :||| :||| :||| :||| :||| :||| :|||
4184 ATTTTGCTTAACGAT...ACCGCAAGCCCCAGGCCAAGACGTCGACAG 4230
17 rArgAlaAspTyrIleuIleLysLeuAsnLysAspLeu...AlaArgL 33
||||| :||| :||| :||| :||| :||| :||| :||| :|||
4231 GCGTCCGAGTACTGCTCAAGATCATCAAGAAAGCTGTGAGTACCA 4280
33 ysgLualaglnArg 37
||| :||| :||| :||| :||| :||| :||| :||| :|||
4281 AGGAGAGCAACACG 4294
```

seq\_name: /SIDSI/gcgdata/hold-geneseq/genesegn-emb1/NA2001B.DAT:ABL06442

seq\_documentation\_block:

ID ABL06442 standard; CDNA: 9933 BP.

AC ABL06442;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.

KW Drosophila: developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ss.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

PA (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.

DR P-PSDB; ABB62339.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

XX genes from Drosophila and for elucidating cell signalling and cell-cell

XX interactions -

XX Claim 1; SEQ ID NO 13808; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

XX capable of detecting 1000 or more genes from Drosophila. The invention is

XX useful in developmental biology and in elucidating cell signalling and

XX cell-cell interactions in higher eukaryotes for the development of

XX insecticides, therapeutics and pharmaceutical drugs. The invention

XX discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA

XX sequences (ABLI01840-ABLI6175) and the encoded proteins

XX (ABBS7737-ABBS72072).

XX The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;

alignment\_scores:                   Quality: 88.00                   Length: 38  
                                   Ratio: 2.839                   Gaps: 2  
                                   Percent Similarity: 81.579           Percent Identity: 55.263

alignment\_block:  
 US-08-973-363-7 x ABL06442 ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933

1 |||leuproaspappproaspplslyspglnalalysglnleuglnh 17  
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||  
 6724 ATTTCCTTAACGAT...ACGGCAAGCCCAAGCCCAAGCAGCTGAGAC 6770

17 rARGAlaspTyleuileyleuAnlysAspleu...AlaArgL 33  
 |||||:::|||||:::|||||:::|||||:::|||||  
 6771 CGGTGCCGAGACTCTCAGATCATCAAGAGAAAGCGAGCTGAGCA 6820

33 ysglAlaGlnArg 37  
 ||:::|||||  
 6821 ACGGAGGACAAACGC 6834

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAS59590

seq\_documentation\_block:  
 ID AAS59590 standard: DNA: 12928 BP.  
 AC AAS59590:  
 XX AAS59590:  
 XX 13-FEB-2002 (first entry)  
 DE Propionibacterium acnes immunogenic protein encoding DNA #85.  
 XX  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
 KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KM dermatological; osteopathic; neuroprotectant; ds.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN W0200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PE 20-APR-2001: 2001WO-US12865.  
 XX  
 PR 21-APR-2000: 2000US-199047P.  
 PR 02-JUN-2000: 2000US-20884P.  
 PR 07-JUL-2000: 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'malsonneuve J, Zhang Y, Jen S, Carter D;  
 PI  
 DR WPI, 2001-616774/71.  
 XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Claim 1: SEQ ID No 85; 1069PP: English.  
 XX  
 XX Sequences AAS59506-AAS59804 represent DNA molecules encoding  
 CC Propionibacterium acnes immunogenic polypeptides. The proteins and their  
 CC associated DNA sequences are used in the treatment, prevention and  
 CC diagnosis of medical conditions caused by P. acnes. The disorders include  
 CC SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and  
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
 CC in infections of bone, joints and the central nervous system, however it  
 CC is particularly involved in the inflammatory lesions associated with acne  
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a

CC patient comprises contacting a sample with a binding agent that binds to  
 CC the proteins of the invention and determining the amount of bound protein  
 CC in the sample. The polypeptides may be used as antigens in the production  
 CC of antibodies specific for P. acnes proteins. These antibodies can be  
 CC used to downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
 CC polypeptides shown in AA058235-AA058404 and AA067601-AA067602.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published-pct-sequences.  
 XX  
 SO Sequence 12928 BP; 2564 A; 4099 C; 4037 G; 2225 T; 3 other:

alignment\_scores:                   Quality: 61.00                   Length: 40  
                                   Ratio: 2.103                   Gaps: 1  
                                   Percent Similarity: 72.500           Percent Identity: 32.500

alignment\_block:  
 US-08-973-363-7 x AAS59590 ..

Align seg 1/1 to: AAS59590 from: 1 to: 12928

3 Proaspappproaspplslyspglnalalysgln.....leuglnh 17  
 |||||:::|||||:::|||||:::|||||:::|||||  
 7295 CCGGAGGCTCCATCCCGCCAGCCGAGCCGAGACGACTTTCAC 7344

17 rARGAlaspTyleuileyleuAnlysAspleuAlaArgLysG 34  
 ||:::|||||:::|||||:::|||||:::|||||  
 7345 GTCATCGAGCGCGACTCGACGATCGCGAAAGATGTTCCTCCGTRAGCA 7394

34 lualGlnArgLeuAlaGly 40  
 ||:::|||||:::|||||  
 7395 CTGGAGAGCGTRAGCGCA 7414

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAS96071

seq\_documentation\_block:  
 ID AAS96071 standard: CDNA: 1908 BP.  
 AC AAS96071:  
 XX AAS96071:  
 XX 26-FEB-2002 (first entry)  
 DE Leishmania antigen 4G2-83 extended DNA.  
 XX  
 XX Leishmaniasis; Leishmania antigen; immunostimulant; protozoacide; IgG23;  
 KM interleukin-15; PCR primer; Lbhsp83; M15; Lt-1; LbeiF44; LmSP9a;  
 KM MAPS-1A; LmgSP1; LmgSP3; LmgSP5; LmgSP8; LmgSP9; LmgSP13; LmgSP19; ss;  
 KW LcgSP1; LcgSP3; LcgSP4; LcgSP8; LcgSP10; IG6-34; IE6-44; 4A5-63; IB11-39;  
 KW ZA10-37; 4G2-83; 4H6-41; 8G3-100.  
 XX  
 OS Leishmania major.  
 XX  
 PN W0200179276-A2.  
 XX  
 PD 25-OCT-2001.  
 XX  
 PE 05-APR-2001: 2001WO-US11254.  
 XX  
 PR 14-APR-2000: 2000US-0551974.  
 PR 05-MAY-2000: 2000US-0565501.  
 PR 14-AUG-2000: 2000US-0639206.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Reed SG, Campos-Neco A, Webb JR, Dillon DC, Skelky YAW, Bhatia A;  
 PI Coler RM, Probst P;  
 PI  
 DR WPI, 2002-061971/08.

```
DR P-PSDB: AAU71861.
XX
XX New isolated Leishmania antigens, useful for prevention, treatment and
PT diagnosis of leishmaniasis, also related nucleic acids for genetic
PT vaccination
XX
XX Claim 5, Page 190, 193pp: English.
XX
XX The invention relates to polypeptides comprising an immunogenic part of a
CC Leishmania antigen. The leishmania polypeptides and their associated DNA
CC sequences, epitopes and fusion proteins are used in the production of
CC compositions used for inducing a protective immune response against
CC leishmaniasis, for prevention and treatment of the disease. The
CC compositions can also be used generally to treat diseases that respond to
CC interleukin-15 stimulation. In addition, the products may contain an
CC immunostimulant. The sequences represent cDNA molecules encoding
CC Leishmania antigens and PCR primers used to amplify DNA of the invention.
XX
XX Sequence 1908 BP; 386 A; 531 C; 627 G; 358 T; 6 other;
SQ
alignment_scores:
Quality: 60.00 Length: 19
Ratio: 3.529 Gaps: 0
Percent Similarity: 89.474 Percent Identity: 52.632
alignment_block:
US-08-973-363-7 x AAS96071 ..
Align seg 1/1 to: AAS96071 from: 1 to: 1908
2 LeupProaspPProaspLysLysProGlnAlaLysGlnLeuGlnThr 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
352 CTTCCTGATGACCTGATGACCTCCCGGACACGCTGACTGCTCG 401
18 gAlaASP 20
|||||
402 AGGTGAC 408
seq_name: /STD1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ30163
seq_documentation_block:
ID AAZ30163 standard; DNA; 34094 BP.
XX
XX AAZ30163;
AC
XX
XX 26-JAN-2000 (first entry)
DT
XX
XX Complete nucleotide sequence of the PAV-3 genome.
DE
XX
XX PAV-3; defective recombinant PAV vector; live recombinant virus;
KW subunit vaccine; nucleic acid immunisation; gene therapy;
KW genetic disease; hemophilia; cystic fibrosis; cancer; viral infection;
KW acquired immune deficiency syndrome; PAV antigen; porcine pathogen; ds.
XX
XX Porcine adenovirus Type 3.
OS
XX
XX WO953047-A2.
PN
XX
XX 21-OCT-1999.
PD
XX
XX 15-APR-1999; 99WO-US08220.
PF
XX
XX 15-APR-1998; 98US-0081882.
PR
XX
XX (UYSA-) UNITV SASKATCHEWAN.
PA
XX
XX Reddy PS, Tikoo SK, Babluk LA;
PI
XX
XX WPI: 1999-620422/53.
DR
XX
XX New nucleic acids from the genome of porcine adenovirus-3, and derived
PT gene therapy vectors, particularly for immunization
```

```
XX
XX Example 2; Fig 1; 87pp: English.
XX
XX The present sequence represents the complete nucleotide sequence of the
CC genome of porcine adenovirus-3 (PAV-3). The specification also describes
CC a defective recombinant PAV vector comprising inverted terminal repeats
CC (ITR), packaging sequences and at least one heterologous nucleotide
CC sequence (II), but lacking E1 functions. The defective vectors replicate
CC inefficiently in cells (other than helper cells) so are unlikely to be
CC immunogenic. Deletion of the E1 (and optionally other regions) increases
CC the size of heterologous insert that can be packaged. The PAV-3
CC polynucleotide sequences are used to produce (recombinant or defective)
CC vectors that can express heterologous proteins, e.g. for making live,
CC recombinant virus or subunit vaccines, for nucleic acid immunisation or
CC for gene therapy (e.g. of genetic diseases such as hemophilia or cystic
CC fibrosis, cancer, or viral infections, including acquired immune
CC deficiency syndrome), also for in vitro expression of recombinant
CC antigens (for antibody production), antisense RNA, ribozymes or
CC therapeutic proteins. They are also used diagnostically to detect PAV
CC antigens and/or nucleic acid. The vectors may be used in human or
CC veterinary medicine, but particularly for expressing protective
CC determinants of porcine pathogens. Regulatory regions may be used to
CC control expression of heterologous genes. Antibodies raised against PAV-3
CC polypeptides can also be used for diagnosis (to detect PAV-specific
XX antigen).
XX
XX Sequence 34094 BP; 6240 A; 11070 C; 10693 G; 6091 T; 0 other;
SQ
alignment_scores:
Quality: 58.50 Length: 32
Ratio: 2.786 Gaps: 1
Percent Similarity: 65.625 Percent Identity: 46.875
alignment_block:
US-08-973-363-7 x AAZ30163/rev ..
Align seg 1/1 to reverse of: AAZ30163 from: 1 to: 34094
5 AspProaspLysLysProGlnAlaLysGlnLeuGlnThrArgAlaAsp 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7211 GATCAGGAGCGCTTCCACACGCCCACTGCGACGCGATCAGGAGTA 7162
21 rLeuileLysLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGln 36
|:|||||:|||||:|||||:|||||:|||||:|||||:
7161 C.....CTCCTCAACAAGAGCTGTGTGAGAGAGAGAGGCCAG 7125
seq_name: /STD1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ39679
seq_documentation_block:
ID AAZ39679 standard; DNA; 2911 BP.
XX
XX AAZ39679;
AC
XX
XX 02-JUL-1999 (first entry)
DT
XX
XX Renal cancer associated gene.
DE
XX
XX Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO9904265-A2.
PN
XX
XX 28-JAN-1999.
PD
XX
XX 15-JUL-1998; 98WO-US14679.
PF
XX
XX 22-JUN-1998; 98US-0102322.
PR
XX
XX 17-JUL-1997; 97US-0896164.
PR
XX
XX 10-OCT-1997; 97US-0061599.
```





**THIS PAGE BLANK (USPTO)**



seq. documentation block: 619 bp mRNA linear EST 18-OCT-2001

LOCUS B1553556 B1553556 CDKN full-length enriched, 16 days neonate thymus Mus

DEFINITION musculus cDNA clone A130024116.3' similar to U10410 Mouse

DNABinding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION B1553556

VERSION B1553556.2 GI:16268254

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 619)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koyota,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okito,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakawa,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M., and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

COMMENT Unpublished (2001)

On Jun 29, 2000 this sequence version replaced gi:8911286.

FEATURES	SOURCE
Location/Qualifiers	
1. 619	
/organism="Mus musculus"	
/db_xref="taxon:10090"	
/clone="A130024L16"	
/clone_lib="RIKEN full-length enriched, 16 days neonate thymus"	
/tissue_type="thymus"	
/dev_stage="16 days neonate"	
/lab_host="DH10B"	
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia	

Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGGAGAGAGAAGATCCACAGACTCTTTTCTTTTTTTTTVVN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 185.0. Second  
strand cDNA was prepared with the primer adaptor of  
sequence [5' GAGGAGAGATTCGAGATTAAATAAFTATCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from lambda

FLC I."

BASE COUNT	194	a	118	c	161	g	146	t
ORIGIN								

alignment\_scores:

Quality:	187.00	Length:	41
Ratio:	4.561	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	87.805

alignment\_block:  
US-08-973-363-7 x BB155356 ..

Align seg 1/1 to: BB155356 From: 1 to: 619

```

1      ILELSPROASPAPSPProaspIysPysPROGlnAlaIySGInleuGIInH 17
489 ATTCTTCAGATGCATCCTGATATAAAAAACCAACAAGCAAAACACTTACAGAC 538

17   targaAlaasPYrLeuIlleLySLeuLeuAnlysAspleAlaAArgLYSG 34
539 CCGTCCAGACTCTACCTCATCCAACCTTAATTGTGGAGACTYTTGCCAAAAGAAG 588
                                     |||||
34   luAlagInArgeLuAlaIyAla 41
589 AGGCTCAGAGACTTGTGTGGTGCG 611

seq_name: gb_est1:AL644594

seq_documentation_block:
LOCUS       AL644594                645 bp    mRNA           linear     EST 07-NOV-2000
DEFINITION  AL644594 XGC-egg Silurana tropicalis cDNA clone lIEId12 5', mRNA
sequence..
ACCESSION   AL644594
VERSION     AL644594.1  GI:16796719
KEYWORDS    EST.

SOURCE      western clawed frog.
ORGANISM    Silurana tropicalis
            Euharyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
            Xenopodinae; Silurana.
REFERENCE   1 (bases 1 to 645)
AUTHORS    Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
TITLE       Sanger xenopus tropicalis EST project 2001 (10_2001)
JOURNAL    Unpublished (2001)
COMMENT     Contact: Huckle E
            Sanger Centre
            Hinxton, Cambridgeshire, CB10 1SA, UK
            Email: tropes@sanger.ac.uk
            Sanger xenopus tropicalis EST project 2001
            TROPICALIS_SEQUENCE_ID: LIEId12.plc
            Sequencing primer: PLC
            This sequence is from a Xenopus gene collection (XGC) library
            constructed by Aaron M. Zorn.

FEATURES             Location/Qualifiers
     source            1..645
                        /organism="Silurana tropicalis"
                        /db_xref="taxon:8364"
                        /clone="LIEId12"

```

```

/clone_1lb="XGC-egg"
/dav_stage="egg"
/label="Escherichia coli XLI-blue"
/Note-Vector: PCS107; site_1: EcoRI; site_2: NotI; cDNA
was oligo dt primed from 5' end of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into PCS107 with
EcoRI at the 5' end and NotI at the 3' end"

BASE COUNT      222 a      125 c      156 g      141 t      1 others

ORIGIN

Alignment Scores:
      Quality: 187.00      Length: 39
      Ratio: 4.921      Gaps: 0
Percent Similarity: 97.436      Percent Identity: 94.872

alignment_block:
US-08-973-363-7 x AL644594

Align seg 1/1 to: AL644594 from: 1 to: 645

1 11eleuProApsaPronApLyLySProGlnAlaLySglnleuGlnTh 17
|||||
456 ATTTTACCAAGATGATCCAGATAGAACCCCAAGCAAGCAAGCATAGAGC 505
|||||
17 TATGAlaSPyTleuIlleLySleuLeuAsnLySAspLeuAlaLySg 34
|||||
506 CAGAGTGACACCTCATTTAAACCTTCATTAAGATCTGTTAGGAAG 555
|||||
34 lulaGlnArgLeuAla 39
|||||
556 AACCAACAAGACTTCT 572

seq_name: gb_csti:BB461065

seq. documentation block:
LOCUS      BB461065      660 bp      mRNA      linear      EST 25-OCT-2001
DEFINITION      BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion
                DMS-musculus cDNA clone D13070513.3 similar to L10410 Mouse
                DNA-binding protein (CNP-1) mRNA, mRNA sequence.
ACCESSION      BB461065
VERSION        BB461065.2      GI:16426612
KEYWORDS
SOURCE        EST.
ORGANISM      house mouse.
                Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
                1 (bases 1 to 660)
                Atsuka,T., Carrinoni,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A
                , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
                ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
                Okazaki,Y., Okitsu,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasak
                ,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
                Tanihara,M., Tagawa,A., Takahashi,T., Takeda,Y., Tanaka,T., Toyota,T.,
                Muramatsu,M. and Hayashizaki,Y.
                RIKEN Mouse ESTs (Arikawa,T., et al. 2001)
                Unpublished (2001)
                On Jul 21, 2000 this sequence version replaced gi.9356558.
TITLE
JOURNAL
COMMENT
Contract: Yoshitake Hayashizaki
Laboratory for genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
url:http://genome.gsc.riken.go.jp/
Carrinoni,P., Shibata,K., Hayashi,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subcloning of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagl,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

```

```

Matsubaki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1457-1471 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 261-269 (2001)
Kondo,S., Shingara,A., Saito,T., Kiyosawa,H., Yamada,I., Alizawa
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
Hayashizaki,Y.
Comptional Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.ysc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers
1..660
/organism="Mus musculus"
/db_xref="taxon:10090"
/cclone="D13007B13"
/clone.lib="RIKEN full-length enriched, 12 days embryo
spinal ganglion"
/tissue-type="spinal ganglion"
/dev-stage="12 days embryo"
/lab_host="DH10B"
/Note="Site_1: Salt; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
plined with a primer [5'
GAGGAGAAGACGGCGCCGCACCTCGAATTTTTTTTTTTTNN 3'] , cDNA was
prepared by using triethanol thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGGAGAAGATTCCTCGAATTAAATTAATTAATWCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
phluescript KS(+) after bulk excision from Lambda PLC I."
BASE COUNT      211 a      130 c      168 g      151 t
ORIGIN

Alignment_scores:
Quality: 187.00 Length: 41
Ratio: 4.561 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 87.805

Alignment block:
US-08-973-363-7 x BB461065 ..

Align seg 1/1 to: BB461065 from: 1 to: 660

1   lileuproaspaspproksplsyisyproglnalysglnleugclnth 17
|||||
461 ATTCTTCGAGATGATCTCATATAAAAACCACAGCAAAACGTTCACAGC 530
|||||
17  rtaglaalspytleutllelyseuleuanlsyaspjeualaalygsg 34
|||||
531 CCGTGAGACTACTCTCATCAACTCTTACGACAGATCTTGCAAAAAGAG 580
|||||
34  luiaaglnargleualaglalyala 41
|||||
581 AGGCTTCAGAGACTTGTGGTGGC 603

```

seq\_name: gb\_est1:BB834922

seq\_documentation\_block:

LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001

DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527 Jy9-MC(B) cDNA Mus musculus cDNA clone G93003J421 3', mRNA sequence.

ACCESSION BB834922

VERSION BB834922.1 GI:17013165

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 446)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gscl.riken.go.jp, URL: http://genome-gsc.riken.go.jp/

COMMENT Carninci,P., Shibata,Y., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

AUTHORS wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Tameda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

TITLE Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site (<http://genome-gsc.riken.go.jp>) for further details.

FEATURES

source location/Qualifiers

1. 446

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="G93003J421"

/clone\_1lb="RIKEN full-length enriched, mammary gland RCB-0527 Jy9-MC(B) cDNA"

/tissue\_type="mammary gland"

/cell\_line="RCB-0527 Jy9-MC(B)"

BASE COUNT 139 a 99 c 108 g 100 t

ORIGIN

alignment\_scores: Quality: 182.00 Length: 41 Ratio: 4.439 Gaps: 0 Percent Similarity: 100.000 Percent Identity: 82.927

alignment\_block:

us-08-973-363-7 x BB834922

Align seg 1/1 to: BB834922 from: 1 to: 446

1 l1eudProaspaspProaspLysProGlnAlaLysGlnLeuGln17  
|||||  
309 ATTCTCCAGATGATCTGATTAACCAAGCCAAACAGGTACGAC 358  
17 rargAlaAspTyrIleuIleuLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
359 CCGTCAGACTACCTCCTCAACTACTTACGACGATCTTGCCAAAGAG 408  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
409 AGGCTCAGACACTTGTGTTTCG 431

seq\_name: gb\_est1:BB830730

seq\_documentation\_block:

LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001

DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527 Jy9-MC(B) cDNA Mus musculus cDNA clone G930013R04 3', mRNA sequence.

ACCESSION BB830730

VERSION BB830730.1 GI:17008973

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 438)

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gscl.riken.go.jp, URL: http://genome-gsc.riken.go.jp/

COMMENT Carninci,P., Shibata,Y., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

AUTHORS wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Tameda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

TITLE Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Please visit our web site (<http://genome-gsc.riken.go.jp>) for further details.

FEATURES

source location/Qualifiers

1. 438

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="G930013R04"

/clone\_1lb="RIKEN full-length enriched, mammary gland RCB-0527 Jy9-MC(B) cDNA"

/tissue\_type="mammary gland"

/cell\_line="RCB-0527 Jy9-MC(B)"

BASE COUNT 139 a 99 c 108 g 100 t

ORIGIN

alignment\_scores: Quality: 182.00 Length: 41 Ratio: 4.439 Gaps: 0 Percent Similarity: 100.000 Percent Identity: 82.927



Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

BASE COUNT 114 a 118 c 85 g 230 t

ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-7 x AI890775/rev ..

Align seg 1/1 to reverse of: AI890775 from: 1 to: 547

1 ILEUProASPpProASPpLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
378 ATTCTTCAGATGATCCGATATAAAACCAAGCAAAACAGTTGCAGAC 329  
17 FARGAlaASPpTyrlLeuLleLysLeuLeuAsnLysASPpLeuAlaArgLysG 34  
|||||  
328 CCGTGCAGACTACTCTCATCAATTACTTAGAGATCTTGCAAAAAAAG 279  
34 lUaGlnArgLeuAlaGlyAla 41  
|||||  
278 AAGCT.....CTTTCGTGTCGCG 262

seq\_name: gb\_est1:AU125712

seq\_documentation\_block:  
LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2M4002061 5', mRNA  
sequence.  
ACCESSION AU125712  
VERSION AU125712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 866)  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.  
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
,Y., Sugano,S., Isogai,T.)  
Unpublished (2000)  
JOURNAL  
COMMENT  
CONTACT: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

FEATURES  
source  
1. 866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_lib="NT2RM4"  
/tissue\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t 2 others

ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-7 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

1 ILEUProASPpProASPpLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
450 ATTCTTCAGATGATCCGATATAAAACCAAGCAAAACAGTTGCAGAC 499  
17 FARGAlaASPpTyrlLeuLleLysLeuLeuAsnLysASPpLeuAlaArgLysG 34  
|||||  
500 CCGTGCAGACTACTCTCATCAATTACTTAGAGATCTTGCAAAAAAAG 549  
34 lUaGlnArgLeuAlaGlyAla 41  
|||||  
550 AAGCT.....CTTTCGTGTCGCG 566

seq\_name: gb\_est2:BE895133

seq\_documentation\_block:  
LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.  
ACCESSION BE895133  
VERSION BE895133.1 GI:10358221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1028)  
NIH-MGC http://mgc.ncl.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
JOURNAL  
COMMENT  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: ATCC/DC/D/DTF  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM9753 row: h column: 16  
High quality sequence stop: 488.  
Location/Qualifiers  
1. 1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Skin; Vector: PCMV-SPOrt6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 387 a 205 c 238 g 198 t

ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:





Align seg 1/1 to: BF239967 from: 1 to: 821

```
6 ProasplylsysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLe 22
|||||
3 CCCGATATTAACCAACAGCAAGCAAGTTCGACAGCCCGTCGAGACTACT 52
|||||
22 uileuProasplylsysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLe 39
|||||
53 CATCAATTAAGTCTAGTAGATCTGCAATTAAGTCTGCAATTAAGTCT 96
|||||
39 laGlyAla 41
|||||
97 CTGCTGCG 104
```

seq\_name: gb\_est1:AW996787

seq\_documentation\_block:

LOCUS AW996787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM996787  
VERSION AM996787.1 GI:8257021

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 337)

AUTHORS

Nagai,M.A., da Silva,W. Jr., Zagzo,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE

COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-230

200-102-d03&tl=2000-02-23&tl=1)

Seq primer: puc 18 forward

High quality sequence start: 2

High quality sequence stop: 337.

Location/Qualifiers

FEATURES

source

1..337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"

/note="Organ: breast.normal; Vector: puc18; Site:1: Sma1;  
Site:2: Sma1: A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

ORIGIN

alignment\_scores:

Quality: 133.50

Ratio: 3.926

Percent Similarity: 94.444

Length: 36

Gaps: 2

Percent Identity: 83.333

alignment\_block:

us-08-973-363-7 x AW996787/rev ..

Align seg 1/1 to reverse of: AM996787 from: 1 to: 337

```
1 lleuProasplylsysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLe 17
|||||
105 ATTCTTCAGATGATCCGATTAACCAACAGCAAGTTCGACAGAC 56
|||||
17 rArgAlaAspTyrLeuileLys.LeuLeuAsnLysAspLeuAlaArgLys 33
|||||
55 CCGTCGACAGACTATCTCATCACTACTT...AGTAGAGATCTTGCAAAAAA 9
|||||
34 GluAla 35
|||||
8 GAGACT 3
```

seq\_name: gb\_est1:AM997058

seq\_documentation\_block:

LOCUS AM997058 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM997058  
VERSION AM997058.1 GI:8257292

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 686)

AUTHORS

Nagai,M.A., da Silva,W. Jr., Zagzo,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE

COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-150

400-152-c03&tl=2000-04-15&tl=1)

Seq primer: puc 18 forward

High quality sequence start: 19

High quality sequence stop: 678.

Location/Qualifiers

FEATURES

source

1..686  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"

/note="Organ: breast.normal; Vector: puc18; Site:1: Sma1;  
Site:2: Sma1: A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 165 a 154 c 126 g 241 t

ORIGIN



---



source 1..6608 /organism="unidentified" /db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN  
alignment\_scores:  
Quality: 205.00 Length: 41  
Ratio: 5.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-8 x A58691 ..  
Align seg 1/1 to: A58691 from: 1 to: 6608  
1 lleleuproaspaprosaplyslysproglinalalysglnleuglnth 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAAGCCAGCAAGAGCTACAGAC 4129  
17 fargalaspptyrleulleulsleuasnlysaspseuAlaArglysg 34  
|||||  
4130 CCGTGCAGACTACTCTATTAAATTACTGATTAAGACCTTGCAAGAAAG 4179  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
4180 AAGCACAAGAGCTTGCTGTGTGCA 4202  
seq\_name: gb\_ov:AF004397  
seq\_documentation\_block:  
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
DEFINITION Gallus chromo-helicase-DNA-binding on the z chromosome  
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete  
cds  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE  
ORGANISM chicken.  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 6872)  
REFERENCE  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE A CHD1 gene is z chromosome linked in the chicken Gallus domesticus  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
REFERENCE 2 (bases 1 to 6872)  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK  
FEATURES  
source  
1..6872  
Location/Qualifiers  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1..6872  
/gene="CHD-Z"  
228..5654  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the z chromosome  
protein"  
/protein\_id="AAC60282.1"  
/translation="MNGHSDSESVNNSGESSRSDSGASGSGSGSSGSSSS  
SOSGSDSESGSGSSESSTREKKQYAKPKADGSEFMKSPETILAVQSAV  
LKRQOQOKRAASDSGSESDSSSDSDSSSTETKKKHDDQMSGSGSVGTGS  
DSEADEGDKSSCESSESDEYEPKRVKSKRPSPRIKPKSGKSTGTGOKRQDLSSEEE

DDDEDYDKRGSRROATVNVSYKEAEFTKSDLLLEVCGEDVPQTEDEFEETIEKFM  
SRIGRGATGASTTIYAAVEDDGPVAGCEKSELEIYLIJLKMKGMSIHMTWEET  
LKOQVKGKMKLDNKKKDOETKRWLKNASPDVYVYCOOELTPDLKQOYVRII  
AHSNOKSAAGYDPDYCKMOGLPYSPCSMDGALIAKKQOARIDEVFSNOSKTPFKD  
CYLQORPFAVALKKOPSTYIGHESELRLDYOLNMLNLAHMKCKGNCLADDEGCG  
KTQITSPFNVLFEHOLGPEPLVPLSTLTSWOREIQVAPQNAVAVLADITSRN  
MIRTHEMHPQTRKLFNLIITTYEILDKDSFGLGMWAFGVDEARLKNDSLY  
RTLIDFKSNHRLITGTPLONSLAKELSLHEIMPEKFSWDEPEEGKRGVYAS  
LAKELEPEFLRRVKOVESLPAKVEQLIRMEASALQOYKWIITRYKALKSGSK  
TSGFLNIMELKCCNHCYILKPPDDNDFYKQALOHILRSSGKLLILKILRL  
PBGNVILFSCWVRMLDILAEYLKRYOPFORLDGISTGELRKAOLDHFNKSGSDPC  
FLSTRAGGLGINTLASDVTYITDSMPONDLOQANRHRIGOKQOYNIIRLYTSG  
VEEDILERRAKKMWLDHLVLOMDTGTGTVLHTGSTPSSSTPFNKEELSAITKFAE  
LEKEPEGEEOEPQMDIDELIKRAETRENEGPLVGBELLQFKNVAFSNDEIDIE  
LEPERNSRWMEELIPEOSORRIEEEROKELEIYMLPRMRCAKOISFNSGERSR  
SRVYSGSDSDSTTERKPRKRGRRPTRENTKGSDAEIRFRTISYKPGPLERLD  
AVARDAELVDKSEPTLRRLGELVHNCITKALKDSSGOBRAGRGKCGPFRISGV  
QVNAKLVISHEBELPLKHSIPSDPEERKRYTPHTAAHEDITMGKEDDSNLVGI  
YEGYGSWEMIIMKDPDLSLTOKILPDDPDKQOALQIOTRADYILIKLNLKDLAREAQ  
RLAGAGNSKRRTKRNKMKASKIREEIKSDSSPOSEKSDDEEDDNKDEIVSVK  
HLKKIKTEKEENEKEPEPDIGIKKEAEERETKEKREKREKEDKELKED  
NKEKRENVKESSTOKKEVKEEKVEMKSENKESKTIPLDTPVHTATSPRPISE  
ESELHOKTFSYCKERMRYKALKQIDOREGSLSEPOLEHTTRCIIKIGHTRECL  
KETNPEQIKQMRKMLTFVSKFTFEDKRLKLYKNAIKRQESQINDQNTSNVN  
THVIRNPVERLKETTNHDDSSRDSYSSDRHLISQYHDDHDKDHQDAVKKSDSRPY  
SAFSNGKDHDDMDHYKODSRYSYSDSKHRRLDHRSDRHSNLEGMKDSRHSRSH  
SDRLHSDHRSSTSEYSHKSSRDYRHSYHSDWQMDHRASSGSPRSLDQSPGSRPLG  
HRSPEHSSDHKSTPEHWTWSKKT"  
misc\_feature  
4341..4604  
/gene="CHD-Z"  
/note="short insert found in longer variant mRNA of CHD-Z"  
BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN  
alignment\_scores:  
Quality: 205.00 Length: 41  
Ratio: 5.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-8 x AF004397 ..  
Align seg 1/1 to: AF004397 from: 1 to: 6872  
1 lleleuproaspaprosaplyslysproglinalalysglnleuglnth 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAAGCCAGCAAGAGCTACAGAC 4129  
17 fargalaspptyrleulleulsleuasnlysaspseuAlaArglysg 34  
|||||  
4130 CCGTGCAGACTACTCTATTAAATTACTGATTAAGACCTTGCAAGAAAG 4179  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
4180 AAGCACAAGAGCTTGCTGTGTGCA 4202  
seq\_name: gb\_pat:A58684  
seq\_documentation\_block:  
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 3 from Patent WO9639505.  
ACCESSION A58684  
VERSION A58684.1 GI:3712427  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tjuri,B.  
TITLE AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;

ISIS INNOVATION (GB)  
COMMENT Other publication AU 5906996 961224.  
FEATURES Location/Qualifiers  
source 1..153  
/db\_xref="taxon:32644"  
BASE COUNT 58 a 40 c 31 g 24 t  
ORIGIN  
alignment\_scores:  
Quality: 190.00 Length: 51  
Ratio: 4.634 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 80.392  
Alignment block:  
US-08-973-363-8 x A58684 ..  
Align seg 1/1 to: A58684 from: 1 to: 153  
1 ATTTACTGATGATTCAGATTAAGAAACCCAGGCAAGACGCTACAGAC 50  
|||||  
1 ATTTACTGATGATTCAGATTAAGAAACCCAGGCAAGACGCTACAGAC 50  
|||||  
8 LysylsProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuLeu 24  
|||||  
51 CAAGAAACCCAGGCAAGACGCTACAGACCCGCTGAGACTACTCTATTA 100  
|||||  
24 ystLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
|||||  
101 AATTACTGATTAAGACCTTCCAGAAAGGAGACCAAGGCTTGCTGGT 150  
|||||  
41 Ala 41  
151 GCA 153  
seq\_name: gb\_pat:A58685  
seq\_documentation\_block:  
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 4 from Patent WO9639505.  
ACCESSION A58685  
VERSION A58685.1 GI:3714248  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
Griffiths, R. and Tjwaari, B.  
AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS  
Patent: WO 9639505-A 4 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Location/Qualifiers  
source 1..153  
/db\_xref="taxon:32644"  
/organism="unidentified"  
BASE COUNT 56 a 36 c 31 g 30 t  
ORIGIN  
alignment\_scores:  
Quality: 190.00 Length: 51  
Ratio: 4.634 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 80.392  
Alignment block:  
US-08-973-363-8 x A58685 ..  
Align seg 1/1 to: A58685 from: 1 to: 153  
1 ATTTACTGATGATTCAGATTAAGAAACCCAGGCAAGACGCTACAGAC 50  
|||||  
1 ATTTACTGATGATTCAGATTAAGAAACCCAGGCAAGACGCTACAGAC 50  
|||||  
8 LysylsProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuLeu 24  
|||||  
51 CAAGAAACCCAGGCAAGACGCTACAGACCCGCTGAGACTACTCTATTA 100  
|||||  
24 ystLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
|||||  
101 AATTACTGATTAAGACCTTCCAGAAAGGAGACCAAGGCTTGCTGGT 150  
|||||  
41 Ala 41  
151 GCA 153  
seq\_name: gb\_ro:MUSCHD1X  
seq\_documentation\_block:  
LOCUS MUSCHD1X 5349 bp mRNA linear ROD 19-SEP-1996  
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.  
ACCESSION L10410 X66028  
VERSION L10410.1 GI:455014  
KEYWORDS  
SOURCE mouse.  
ORGANISM mouse musculus.  
REFERENCE 1 (bases 1 to 5349)  
Delmas, V., Stokes, D.G., and Perry, R.P.  
A mammalian DNA-binding protein that contains a chromodomain and an  
SNE2/SWI2-like helicase domain  
Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
93211972  
JOURNAL  
MEDLINE  
REFERENCE 2 (bases 1 to 5349)  
Perry, R.P.  
TITLE Direct Submission  
AUTHORS Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center,  
7701 Burholme Avenue, Philadelphia, PA 19111, USA  
COMMENT On Feb 16, 1994 this sequence version replaced gi:293322.  
FEATURES Location/Qualifiers  
source 1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/codon\_start=1  
/product="DNA-binding protein"  
/protein\_id="FAB08486.1"  
/db\_xref="GI:455015"  
/translation="MNGSDPESVRNGSGESGSDGCGSASGSGSGSSGSSDSS  
SGSGSSDSDSGSDSGSGSSESDTSREKRYQKPKVDGKPKSSPSLAVQSAHL  
RQKQDQKQKSSSSGSSSESDSSSSGSKKRNKNDQKSGSSGSGSDG  
SSGSEBERKSSCDGTSSTLEPKKYSKSKQKSKKSKKTLQKROKQSDSDG  
DDEDLDNKRSSKQDTYVSTKDEBKATDDDLAEKCEDEYPODEBERETTERM  
DKRGKRGATGATTTTAVADGSPNAGTERKKERGDQYDLKMKMSKSHINTEETE  
THQKQVSGKAKLDNKKRQDTERKMLKNAKSPEDVEITIKQDELTDLKHQIQVERE  
TAKQVKSAGGLPDYCKMGGLPTECSWEDGLSKKFTCIDETFSRNSKTPK  
DKYVKQRPYVALKKQPSYVIGHELELDLQNLGKLMHAWKSGNSCLADGINSR  
GNTIOTISPLNTLFEHOLYGPFLVPLSTLTSWQRIOTWASQKNAVVLGINSR  
NMTFHEMRHPQTRKRLKNTLTTEYLILKDAFLGGLNAFLGVDEHARKNDSSL  
KXTLIDFKSNHRLITGPIQNSLKELEWLSLHFTMPKFSWEDPEEGRGREGYA  
SLHKELEPLLRVKKDKVKSIPAKVQELIREMSALQKQYKMLITRYKALSLSK  
GSTGSPNIMWELKCCNHCEYLKPPDNEFYKQDALQYKMLITRYKALSLSK  
REKGRVLIIFSQWVRMDIILAELYKROFPQRLDGSIKELRKQALDHFPAESDEP  
CEFLSTRAGGLGINLASADYVIGSDMNPONDLOAARHARGOKAVNTLRYLVTG  
SVEEDILERRAKKKMYDHLVIQRMDTGTGTVLHNGSAPSSPTFNKEELASILKFGAE  
ELFKPEGGEQEPQMDIDELIKRAETHENEGPLSYGDELILQFVYANFSNMDEDI  
ELPEPNSKNWEDILPEQORRLLEERQKLELEIYMLPMRCAKQISFNGSGRS  
RSRRSGSDSDSISERKPKRGRGRTPTIPRENIKGFSDAIRRFISKYKKGGLERL  
DALRADAEIVDKSETDRLRLGELVHNCVAKLKSSSOTERAGRGKVGKQPTPTISG

|||||  
1 ATTTACTGATGATTCAGATTAAGAAACCCAGGCAAGACGCTACAGAC 50  
|||||  
8 LysylsProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuLeu 24  
|||||  
51 CAAGAAACCCAGGCAAGACGCTACAGACCCGCTGAGACTACTCTATTA 100  
|||||  
24 ystLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
|||||  
101 AATTACTGATTAAGACCTTCCAGAAAGGAGACCAAGGCTTGCTGGT 150  
|||||  
41 Ala 41  
151 GCA 153  
seq\_name: gb\_ro:MUSCHD1X  
seq\_documentation\_block:  
LOCUS MUSCHD1X 5349 bp mRNA linear ROD 19-SEP-1996  
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.  
ACCESSION L10410 X66028  
VERSION L10410.1 GI:455014  
KEYWORDS  
SOURCE mouse.  
ORGANISM mouse musculus.  
REFERENCE 1 (bases 1 to 5349)  
Delmas, V., Stokes, D.G., and Perry, R.P.  
A mammalian DNA-binding protein that contains a chromodomain and an  
SNE2/SWI2-like helicase domain  
Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
93211972  
JOURNAL  
MEDLINE  
REFERENCE 2 (bases 1 to 5349)  
Perry, R.P.  
TITLE Direct Submission  
AUTHORS Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center,  
7701 Burholme Avenue, Philadelphia, PA 19111, USA  
COMMENT On Feb 16, 1994 this sequence version replaced gi:293322.  
FEATURES Location/Qualifiers  
source 1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/codon\_start=1  
/product="DNA-binding protein"  
/protein\_id="FAB08486.1"  
/db\_xref="GI:455015"  
/translation="MNGSDPESVRNGSGESGSDGCGSASGSGSGSSGSSDSS  
SGSGSSDSDSGSDSGSGSSESDTSREKRYQKPKVDGKPKSSPSLAVQSAHL  
RQKQDQKQKSSSSGSSSESDSSSSGSKKRNKNDQKSGSSGSGSDG  
SSGSEBERKSSCDGTSSTLEPKKYSKSKQKSKKSKKTLQKROKQSDSDG  
DDEDLDNKRSSKQDTYVSTKDEBKATDDDLAEKCEDEYPODEBERETTERM  
DKRGKRGATGATTTTAVADGSPNAGTERKKERGDQYDLKMKMSKSHINTEETE  
THQKQVSGKAKLDNKKRQDTERKMLKNAKSPEDVEITIKQDELTDLKHQIQVERE  
TAKQVKSAGGLPDYCKMGGLPTECSWEDGLSKKFTCIDETFSRNSKTPK  
DKYVKQRPYVALKKQPSYVIGHELELDLQNLGKLMHAWKSGNSCLADGINSR  
GNTIOTISPLNTLFEHOLYGPFLVPLSTLTSWQRIOTWASQKNAVVLGINSR  
NMTFHEMRHPQTRKRLKNTLTTEYLILKDAFLGGLNAFLGVDEHARKNDSSL  
KXTLIDFKSNHRLITGPIQNSLKELEWLSLHFTMPKFSWEDPEEGRGREGYA  
SLHKELEPLLRVKKDKVKSIPAKVQELIREMSALQKQYKMLITRYKALSLSK  
GSTGSPNIMWELKCCNHCEYLKPPDNEFYKQDALQYKMLITRYKALSLSK  
REKGRVLIIFSQWVRMDIILAELYKROFPQRLDGSIKELRKQALDHFPAESDEP  
CEFLSTRAGGLGINLASADYVIGSDMNPONDLOAARHARGOKAVNTLRYLVTG  
SVEEDILERRAKKKMYDHLVIQRMDTGTGTVLHNGSAPSSPTFNKEELASILKFGAE  
ELFKPEGGEQEPQMDIDELIKRAETHENEGPLSYGDELILQFVYANFSNMDEDI  
ELPEPNSKNWEDILPEQORRLLEERQKLELEIYMLPMRCAKQISFNGSGRS  
RSRRSGSDSDSISERKPKRGRGRTPTIPRENIKGFSDAIRRFISKYKKGGLERL  
DALRADAEIVDKSETDRLRLGELVHNCVAKLKSSSOTERAGRGKVGKQPTPTISG

VOVNAKLVIAHEDELIPLHKSTPSDPERKQYITPCHTKAHFIDWGENEDSNLLIG  
IYEVYGSWEMITKMDPDLSTJTKILPDDPKRPAKOLQTRADYILIKLSRLAKREA  
ORICGAGSMETRRKTRAKSKAMKSIKKEIKDSKSPLESEKSDDDDKLNDREKPEK  
DESKSVSDAPVHTTASGEVPIAESEPELOKFEJCKEMRKYKAKALOLDPEK  
GLSEPEOLEHTPROCLIKIGDHTTECLKEYSNEOLKOWRKNIMIVSKPTEPDARKH  
KLXKAIAKKRQESQONSNDONSNAVTHYIRNDEMRLKENTVHDDSSRDSYSDDRLS  
QTHDHKDHQGDYKSKSDSRKPYSSFSNGKDHREWDHYRODSRYSDREXHKRLD  
HRSRHRPSLEGLDRCHSDHRSYSDHMSHDHSSSEHTHKSSRDYRYLSDWOLD  
HRAASSGPRSPIDORSPYGSRSPFESHAENRSTPEHTWSSRKT"

BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN

alignment\_scores:  
Quality: 187.00 Length: 41  
Ratio: 4.561 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.805

alignment\_block:  
US-08-973-363-8 x MUSCHDX ..

Align seg 1/1 to: MUSCHDX from: 1 to: 5349

1 lleuProaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
4026 ATTCTTCACAGATGATCCTGATATAAACCAACCAACAGTACACAG 4075  
|||||  
17 ratGAlaSPtyrLeuLeuLeuLeuAnLysAspLeuAlaArgLysG 34  
|||||  
4076 CCGTGCAGACTACCTCATCAACTACTTAGCAGATCTTGCAAAAGAG 4125  
|||||  
34 lnaGlnArgLeuAlaGlyAla 41  
|||||  
4126 AGGCTCAGAGACTTGTGTGCGG 4148  
|||||

seq\_name: gb\_pat:A58686

seq\_documentation\_block:  
LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 5 from Patent WO9639505.  
ACCESSION A58686  
VERSION A58686.1 GI:3714249  
KEYWORDS  
SOURCE  
ORGANISM  
unclassified.  
REFERENCE  
1 (bases 1 to 153)  
AUTHORS Griffiths R. and Tiwari B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES  
source location/Qualifiers  
1..153  
/organism="unclassified"  
/db\_xref="taxon:32644"

BASE COUNT 60 a 34 c 31 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 182.00 Length: 51  
Ratio: 4.439 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 76.471

alignment\_block:  
US-08-973-363-8 x A58686 ..

Align seg 1/1 to: A58686 from: 1 to: 153

1 lleuProaspProasp..... 7  
|||||

1 ATTTACCTGATGACCGAGATAGAAACACAGGCAAGAGTTGCAGAC 50  
8 LysLysProGlnAlaLysGlnLeuGlnThArgAlaSPtyrLeuLeu 24  
|||||  
51 CAAAGAACACAGGCAAGAGAGTTCAGACCCGTCGAGATTACTCATTTA 100  
|||||  
24 ystLeuLeuAnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
|||||  
101 AATTACTGAAATAAAGACCTTGCAAGAAAGAGTCAAGACTTACTGT 150  
|||||  
41 Ala 41  
151 GCA 153

seq\_name: gb\_pat:A58683  
seq\_documentation\_block:  
LOCUS A58683 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 2 from Patent WO9639505.  
ACCESSION A58683  
VERSION A58683.1 GI:3714246  
KEYWORDS  
SOURCE  
ORGANISM  
unclassified.  
REFERENCE  
1 (bases 1 to 153)  
AUTHORS Griffiths R. and Tiwari B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 2 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES  
source location/Qualifiers  
1..153  
/organism="unclassified"  
/db\_xref="taxon:32644"

BASE COUNT 61 a 37 c 27 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 172.00 Length: 51  
Ratio: 4.195 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 70.588

alignment\_block:  
US-08-973-363-8 x A58683 ..

Align seg 1/1 to: A58683 from: 1 to: 153

1 lleuProaspProasp..... 7  
|||||  
1 ATTCTTCACAGATGATCCTGATATAAACCAACCAACAGTACACAG 50  
|||||  
8 LysLysProGlnAlaLysGlnLeuGlnThArgAlaSPtyrLeuLeu 24  
|||||  
51 CAAAGAACACAGGCAAGAGAGTTCAGACCCGTCGAGATTACTCATTTA 100  
|||||  
24 ystLeuLeuAnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
|||||  
101 AACTACTTACAGAGATCTTGCAAGAAAGAGAGCTCAGACACTTGTGT 150  
|||||

41 Ala 41  
151 GCG 153

seq\_name: gb\_pr:AF006513  
seq\_documentation\_block:  
LOCUS AF006513 5947 bp mRNA linear PRI 27-NOV-1997  
DEFINITION Homo sapiens CHD1 mRNA, complete cds.  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428





```

17  TATGALAAAPTYTLEUILEYLSLEUAASNLYSASPLEUALAARGLYSG 34
|||||.....:|||||
26882 CCCTGCGAGTACTCATCAATTAAGTACTAGATCTTGCAAAAAAAG 26833
34  TAAAGT.....CTTCTGCGTGGC 26816
|||||.....:|||||
seq_name: gb_pr:AC012624

seq_documentation_block:
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:14993679
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 134365)
TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Unpublished
2 (bases 1 to 134365)
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134365)
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 134365)
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
location/Qualifiers
source 1. 134365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN

alignment_scores:
Quality: 170.00 Length: 41
Ratio: 4.359 Gaps: 1
Percent Similarity: 95.122 Percent Identity: 85.366

alignment_block:
US-08-973-363-8 x AC012624 ..
Align seg 1/1 to: AC012624 from: 1 to: 134365

1 TLEUAPROASPPROASPLYSLYSPROGLAALALYSGINLEUGINTH 17
|||||.....:|||||
118247 ATCTTCATGATATCCGATTAATAAACCAACCAACAGTTGCAGAC 118296
17 TATGALAAAPTYTLEUILEYLSLEUAASNLYSASPLEUALAARGLYSG 34
|||||.....:|||||
118297 CCGGCGAGTACTCATCAATTAAGTACTAGATCTTGCAAAAAAAG 118346
34 TAAAGT.....CTTCTGCGTGGC 118363
seq_name: gb_htg:AC021449

```

```

seq_documentation_block:
LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION AC021449
VERSION AC021449.3 GI:10047806
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 143079)
TITLE Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Homo sapiens, clone RP11-58M12
Unpublished
2 (bases 1 to 143079)
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2000 this sequence version replaced gi:7407963.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
COMMENT
TITLE JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L5154
Center clone name: 58_M12

----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 134743 bases at least Q40
Consensus quality: 139227 bases at least Q30
Consensus quality: 140814 bases at least Q20
Insert size: 144000; agarose-ip
Insert size: 142179; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-ip
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 38820: contig of 38820 bp in length
* 38821 38920: gap of 100 bp
* 38921 40411: contig of 1491 bp in length
* 40412 40511: gap of 100 bp

```

```

seq.document.location_block: 145659 bp DNA linear HTG 14-FEB-2001
LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE.
VERSION 7 ordered pieces.
ACCESSION AC008531
KEYWORDS HTG: HTSC_PHASE2; HTSC_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 145659)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145659)
AUTHORS DOE Joint Genome Institute.
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
SUBMITTED (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
ON Feb 14, 2001 this sequence version replaced g1:7528342.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 369535
Center clone name: CIT-HSPC_480B11
-----
Summary Statistics
Consensus quality: 139128 bases at least Q40
Consensus quality: 142556 bases at least Q30
Consensus quality: 143744 bases at least Q20
Estimated insert size: 148000; pulse field gel estimation
Estimated insert size: 145059; sum-of-contigs estimation
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
consists of 7 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 56174: contig of 56174 bp in length
56175 56274: gap of unknown length
56275 100874: contig of 44600 bp in length
100875 100974: gap of unknown length
100975 113127: contig of 12153 bp in length
113128 113227: gap of unknown length
113227 118190: contig of 4963 bp in length
118191 118290: gap of unknown length
118291 118694: contig of 1404 bp in length
118695 119794: gap of unknown length
119795 123297: contig of 3503 bp in length
123298 123397: gap of unknown length
123398 145659: contig of 22252 bp in length.
*
FEATURES
source
1. 145659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="45"
/clone="CTC-480B11"
/clone_id="caltech human BAC library C"
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others
ORIGIN
alignment_scores: 170.00 Length: 41
Quality:

```

Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366  
alignment\_block:  
US-08-973-363-8 x AC008531 ..  
Align seg 1/1 to: AC008531 from: 1 to: 145659  
1 11leleProspaspproAspLysPpGlnAlaLysGlnLeuGlnTh 17  
|||||  
46770 ATTCTTCGACATGATCCCGATTAACCAACCAACAAACAGTTGCAGAC 46819  
17 TArGAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
46820 CCGTGACAGACTACCTCATCAATTAAGTAGAGATCTTGCAAAAAAAG 46869  
34 1uA1aGlnAArgLeuAlaGlyAla 41  
|||||  
46870 AAGCT.....CTTCTGTGTGCG 46886  
seq\_name: gb\_hc9:AC091946  
seq\_documentation\_block:  
LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, 33 unordered pieces.  
ACCESSION AC091946  
VERSION AC091946.1 GI:14333882  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 5  
Unpublished  
2 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 544799  
Center clone name: RPCI-11\_36012  
-----  
Summary Statistics  
Consensus quality: 157767 bases at least Q40  
Consensus quality: 180259 bases at least Q30  
Consensus quality: 184175 bases at least Q20  
Estimated insert size: 204590; agarose-fp estimation  
Estimated insert size: 190246; sum-of-contigs estimation  
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation  
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1358: contig of 1358 bp in length  
\* 1359 1458: gap of unknown length  
\* 1459 2667: contig of 1209 bp in length  
\* 2668 2767: gap of unknown length  
\* 2768 3823: contig of 1056 bp in length  
\* 3824 3923: gap of unknown length

3924 5920: contig of 1997 bp in length  
\* 5921 6020: gap of unknown length  
\* 6021 7494: contig of 1474 bp in length  
\* 7495 7596: gap of unknown length  
\* 7597 9096: contig of 1502 bp in length  
\* 9097 9197: gap of unknown length  
\* 9197 10260: contig of 1064 bp in length  
\* 10261 10361: gap of unknown length  
\* 10361 12460: contig of 2100 bp in length  
\* 12461 12560: gap of unknown length  
\* 12561 14611: contig of 2051 bp in length  
\* 14612 14711: gap of unknown length  
\* 14711 16381: contig of 1670 bp in length  
\* 16382 16482: gap of unknown length  
\* 16482 17968: contig of 1487 bp in length  
\* 17969 18068: gap of unknown length  
\* 18069 20434: contig of 2366 bp in length  
\* 20435 20535: gap of unknown length  
\* 20535 23515: contig of 2981 bp in length  
\* 23516 23616: gap of unknown length  
\* 23616 27563: contig of 3948 bp in length  
\* 27564 27664: gap of unknown length  
\* 27664 30987: contig of 3324 bp in length  
\* 30988 31087: gap of unknown length  
\* 31088 36065: contig of 4978 bp in length  
\* 36066 36165: gap of unknown length  
\* 36166 40978: contig of 4813 bp in length  
\* 40979 41078: gap of unknown length  
\* 41079 45663: contig of 4455 bp in length  
\* 45664 45763: gap of unknown length  
\* 45764 51745: contig of 5982 bp in length  
\* 51746 51845: gap of unknown length  
\* 51846 57359: contig of 5514 bp in length  
\* 57360 57459: gap of unknown length  
\* 57460 67881: contig of 10422 bp in length  
\* 67882 67982: gap of unknown length  
\* 67982 74133: contig of 6151 bp in length  
\* 74133 74232: gap of unknown length  
\* 74233 79795: contig of 5563 bp in length  
\* 79796 79895: gap of unknown length  
\* 79896 87511: contig of 7616 bp in length  
\* 87512 87611: gap of unknown length  
\* 87612 92791: contig of 5180 bp in length  
\* 92792 92891: gap of unknown length  
\* 92892 102794: contig of 9903 bp in length  
\* 102795 102894: gap of unknown length  
\* 102895 110866: contig of 7972 bp in length  
\* 110867 110966: gap of unknown length  
\* 110967 117571: contig of 6605 bp in length  
\* 117572 117672: gap of unknown length  
\* 117672 123738: contig of 6067 bp in length  
\* 123739 123838: gap of unknown length  
\* 123839 130583: contig of 6745 bp in length  
\* 130584 130683: gap of unknown length  
\* 130684 141544: contig of 10861 bp in length  
\* 141545 141644: gap of unknown length  
\* 141645 169109: contig of 27465 bp in length  
\* 169110 169209: gap of unknown length  
\* 169210 193446: contig of 24237 bp in length.  
\* Location/Qualifiers  
1. 193446  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-36012"  
/clone\_lib="RPCI human BAC library 11"  
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others  
ORIGIN  
alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1

Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:

US-08-973-363-8 x AC091946

Align seg 1/1 to: AC091946 from: 1 to: 193446

1 lleleuproaspaspProaspIysIysProGlnAlaIysGlnLeuGlnth 17  
 |||  
 87658 ATTCTTCAGATGATCCGATTAACCAACAGCAACAGCAAGCTTGCAGAC 87707

17 rArGAlaAspIyrrleuIlleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
 |||  
 87708 CCGTCGACGACTACCTCATCAATTAAGTACAGATCTTCACAAAAAAG 87757

34 luAlaGlnArgLeuAlaGlyAla 41  
 |||  
 87758 AAGCT.....CTTCTGCTGCG 87774

seq\_name: gb\_pr:AC026778

seq\_documentation\_block:

LOCUS AC026778 195433 bp DNA linear PRI 01-JUN-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.

ACCESSION AC026778  
 VERSION AC026778.4 GI:14277282

KEYWORDS HNG.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 195433)

DOE Joint Genome Institute and Stanford Human Genome Center.  
 Direct Submission

Unpublished  
 2 (bases 1 to 195433)

DOE Joint Genome Institute.  
 Direct Submission

Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 195433)  
 DOE Joint Genome Institute and Stanford Human Genome Center.

Direct Submission  
 Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA

On Jun 1, 2001 this sequence version replaced g1:13677045.  
 Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.9% of Sequence;

Estimated Total Number of Errors is 0.2.  
 STS Content:

WI-13675 G23101  
 SHGC-58345 G38487

SHGC-103595 G57841.  
 Location/Qualifiers

1..195433  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTC-428111"

BASE COUNT 62762 a 37302 c 37040 g 58329 t

ORIGIN

alignment\_scores:  
 Quality: 170.00 Length: 41  
 Ratio: 4.359 Gaps: 1

Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
 US-08-973-363-8 x AC026778/rev

Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433

1 lleleuproaspaspProaspIysIysProGlnAlaIysGlnLeuGlnth 17  
 |||  
 29732 ATTCTTCAGATGATCCGATTAACCAACAGCAACAGCAAGCTTGCAGAC 29683

17 rArGAlaAspIyrrleuIlleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
 |||  
 29682 CCGTCGACGACTACCTCATCAATTAAGTACAGATCTTCACAAAAAAG 29633

34 luAlaGlnArgLeuAlaGlyAla 41  
 |||  
 29632 AAGCT.....CTTCTGCTGCG 29616

---

OM of: US-08-973-363-8 to: N.Geneseq\_032802.\* out\_format: pfs

Date: Aug 3, 2002 4:34 AM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

# Command line parameters:

-MODEL=frame-p2n.model -DEV=x1h  
-Q/cgnt2\_1/USPFO.spool/US08973363/runatc.01082002.080050.18606/app\_query.fasta.1.638  
-DB=N.Geneseq\_032802 -OPMT=fastap -SUFFIX=p2n.rng -GAPOP=12.000  
-GAEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-FGAPOP=4.500 -OGAPOP=0.050 -XGAPOP=10.000 -XGAEXT=0.500  
-FGAPOP=6.000 -FGAEXT=7.000 -YGAOP=10.000 -YGAEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62  
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pcr  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US08973363 -GCCNL\_1.186 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV\_TIMECUT=120 -MARN\_TIMECUT=30 -NO\_XLPPX -WAIT -THREADS=1

# Search information block:

Query: US-08-973-363-8  
Query length: 41  
Database: N.Geneseq\_032802.\*  
Database sequences: 1736436  
Database length: 858457221  
Search time (sec): 523.620000

# Score list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42754 +		205.00	553.09	1.9e-22	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42751 +		205.00	537.11	1.4e-21	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42751 +		205.00	531.94	2.8e-21	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42751 +		190.00	531.94	2.8e-21	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42751 +		182.00	509.45	5.0e-20	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42751 +		172.00	481.33	1.5e-18	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42751 +		172.00	333.38	1.5e-18	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT86882 +		122.00	330.84	4.4e-17	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT86882 +		122.00	330.84	4.4e-17	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLT66443 +		88.00	208.82	0.0028	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLT66443 +		88.00	208.82	0.0028	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABLT66443 +		61.00	123.77	117.17	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAT596071 +		60.00	141.72	15.15	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAAT30163 +		57.50	109.23	977.07	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAAT30163 +		57.50	130.55	63.48	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT95005 +		57.00	130.90	60.67	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAT86057 +		56.50	142.10	14.43	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAT86057 +		56.50	140.09	18.67	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT863085 +		56.50	127.73	91.07	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT863085 +		56.50	127.73	91.07	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAAT75531 +		56.00	127.30	96.34	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAAT75531 +		56.00	127.30	96.34	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAT37101 +		55.00	123.55	155.71	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT53433 +		55.00	134.90	36.33	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT52093 +		55.00	133.58	43.06	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT868525 +		54.50	125.08	127.98	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAT818597 +		54.50	122.06	188.53	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT818597 +		54.50	118.31	304.92	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT818597 +		54.50	112.28	66.01	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT818597 +		54.00	140.29	18.19	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAAT98900 +		54.00	137.71	25.33	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT86834 +		54.00	130.72	61.75	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT86834 +		54.00	123.22	162.46	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT90036 +		54.00	114.85	475.48	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT90036 +		54.00	113.74	548.43	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT814118 +		54.00	107.22	1.3e+0	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT814118 +		54.00	103.61	2.0e+0	
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAAT50904 +		54.00	103.07	2.2e+03	

/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT90033 - 54.00 96.59 4.9  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAAT90032 + 54.00 94.37 6.6  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT86431 + 54.00 73.75 9.1  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT86431 + 53.00 131.84 53.  
/SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAAT86431 + 53.00 128.52 82.

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42754

seq\_documentation\_block:

ID AAAT42754 standard; cDNA: 1316 BP.

AC AAAT42754;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-W gene (partial sequence).

KW Bird; sex determination: chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tlward B;

DR WPT: 1997-043127/04.

PT Avian chromodomain-helicase-DNA binding genes determine sex in

birds - used for sex determin. and to control sex of progeny

Claim 1; Fig 8; 76pp; English.

CC The chicken CHD-W gene (AAAT42754) acting alone or in conjunction with

the closely related CHD-1A gene (AAAT2751) is suggested to initiate

female development in birds. The sequence of CHD-W was deduced

from 2 clones isolated from a 10-day chick embryo library using

a fragment of the CHD-1A gene as a probe. The CHD-W gene is

located on the W chromosome. Probes based on CHD-W and CHD-1A give

a W chromosome-specific signal on hybridisation to genomic DNA of a

non-fertile bird and can be used for sex determin. of a bird. CHD-W

nucleic acids can also be used to control the sex of progeny of a

bird.

Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

alignment\_scores:

Quality: 205.00 Length: 41

Ratio: 5.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignement\_block:

US-08-973-363-8 x AAAT42754

Align seg 1/1 to: AAAT42754 from: 1 to: 1316

1 TleleuProAspAspProAspPylsProGlnAlaIalysGlnleuGlnh 17

1138 ATTTTACTGATGATCCATTAAGAACCCACGCTTACGACG 1187

17 TATGTAAspPylsleuIleuIleuAsnAspAspAspAspAspAsp 34

1188 TCGTGCATATTACTCTATTAAATTTCGATATAAGACCTTCGAGAAAG 1237





```

|||||
1 ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGACTACAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIle 24
51 CAAGAAACCCAGGCAAGACTACAGACCCGTCGACACTTACTCATTA 100
24 yslLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40
101 ATTTACTGATATTAAGACCTTGACAGAAAGAGACAAAGGCTTGCTG 150
41 Ala 41
|||
151 GCA 153

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758
seq_documentation_block:
ID AA142758 standard; DNA; 153 BP.
XX
AC AA142758;
XX
DT 12-MAR-1997 (first entry)
XX
DE Chick CHD-W gene fragment.
XX
KM Bird; sex determination: Chromodomain-Helicase-DNA binding 1;
XX CHD-1A; CHD-W; W chromosome; ss.
XX
OS Gallus sp.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
XX WO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX PA
XX PI Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX
XX P-PSDB: AAM08148.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX PS Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAM08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,
XX foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment_scores:
Quality: 190.00 Length: 51
Ratio: 4.634 Gaps: 1

```

```

Percent Similarity: 80.392 Percent Identity: 80.392
alignment_block:
us-08-973-363-8 x AA142758 ..
Align seq 1/1 to: AA142758 from: 1 to: 153

1 IleLeuProAspAspProAsp..... 7
1 ATTTTACCTGATGATCCAGATTAAGAAACCCAGGCTTAAGCACTTAACAGAC 50
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIle 24
51 CAAGAAACCCAGGCTTAAGCACTTAAGACCCGTCGACACTTACTCATTA 100
24 yslLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40
101 ATTTACTGATATTAAGACCTTGACAGAAAGAGACAGAGACTTGCTG 150
41 Ala 41
|||
151 GCA 153

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759
seq_documentation_block:
ID AA142759 standard; DNA; 153 BP.
XX
AC AA142759;
XX
DT 12-MAR-1997 (first entry)
XX
DE Great tit CHD-W gene fragment.
XX
KM Bird; sex determination: chromodomain-Helicase-DNA binding 1;
XX CHD-1A; CHD-W; W chromosome; ss.
XX
XX Parus major.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
XX WO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX PA
XX PI Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX
XX P-PSDB: AAM08149.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX PS Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAM08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,

```

CC foetus etc. and to manipulate the sex of progeny.  
XX  
SQ Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;  
alignment\_scores:  
Quality: 182.00 Length: 51  
Ratio: 4.439 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 76.471  
alignment\_block:  
US-08-973-363-8 x AAT42759 ..  
Align seg 1/1 to: AAT42759 from: 1 to: 153  
1 lIleuProAspAspProasp..... 7  
1 ATTTTACCTGATGATCCAGATAGAAACACACAGCAAGCATTGACAGAC 50  
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24  
|||||  
51 CAGAAACACACAGCAAGCATTGACAGACCGCTGCAGATTACTCATTA 100  
24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
|||||  
101 AATTACTGAAATAAAGACCTTGCAAGAAAGAAAGTCAAGACTTACTGT 150  
41 Ala 41  
|||  
151 GCA 153  
seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAT42756  
seq\_documentation\_block:  
ID AAT42756 standard; DNA; 153 BP.  
XX  
AC AAT42756;  
XX  
DT 12-MAR-1997 (first entry)  
XX  
DE Mouse CHD-1 gene (bases 3855-9777).  
XX  
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
KM CHD-1; CHD-W; W chromosome; ss.  
XX  
OS Mus sp.  
XX  
FH Key location/Qualifiers  
FT misc\_difference 52..81  
FT /tag= a  
FT /note= "bases 52-81 are a repeat of bases 22-51  
FT and are ignored in the translated amino  
FT acid sequence given in Fig 3"  
XX  
PN WO9639505-A1.  
XX  
PD 12-DEC-1996.  
XX  
PF 05-JUN-1996; 96WO-GB01341.  
XX  
PR 06-JUN-1995; 95GB-0011439.  
XX  
XX (ISIS-) ISIS INNOVATION LTD.  
XX  
XX Griffiths R, Tiwari B;  
XX  
XX WPI, 1997-043127/04.  
XX  
XX P-PSDB; AAM08146.  
XX  
XX Avian chromodomain-helicase-DNA binding genes determine sex in  
XX birds - used for sex determ. and to control sex of progeny  
XX  
XX Claim 8; Fig 3; 76pp; English.  
XX

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
CC and AAT42757). Chicken CHD-W (W refers to the W chromosome) gene  
CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
CC Translated amino acid sequences of this region are provided in  
CC AAM08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
CC AAT42754-55) genes determine sex in birds and can be used to identify  
CC the sex of an embryo, foetus etc. and to manipulate the sex of  
CC progeny.  
XX  
SQ Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;  
alignment\_scores:  
Quality: 172.00 Length: 51  
Ratio: 4.195 Gaps: 1  
Percent Similarity: 80.392 Percent Identity: 70.588  
alignment\_block:  
US-08-973-363-8 x AAT42756 ..  
Align seg 1/1 to: AAT42756 from: 1 to: 153  
1 lIleuProAspAspProasp..... 7  
1 ATTTTACCTGATGATCCTGATAAAAACACACACAAAGCATTACAGAC 50  
8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24  
|||||  
51 CAAAAACACACAGCAAGCATTACAGACCGCTGCAGACTACTCATTA 100  
24 yslLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGly 40  
|||||  
101 AACTACTTAGCAGAGATCTTGCAAAAAAGAGAGCTGACAGACTTGTGT 150  
41 Ala 41  
|||  
151 GCG 153  
seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV59280  
seq\_documentation\_block:  
ID AAV59280 standard; cDNA; 1311 BP.  
XX  
AC AAV59280;  
XX  
DT 14-DEC-1998 (first entry)  
XX  
DE Altered telomere repeat binding factor 1 gene.  
XX  
XX ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;  
KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.  
XX  
XX Homo sapiens.  
XX  
XX Synthetic.  
XX  
FH Key location/Qualifiers  
FT CDS 1..1311  
FT /tag= a  
FT /product= "A-TRF"  
XX  
XX WO9836066-A1.  
XX  
XX 20-AUG-1998.  
XX  
XX 13-FEB-1998; 98WO-US02765.  
XX  
XX 04-FEB-1998; 98US-0018628.  
XX  
XX 13-FEB-1997; 97US-0800264.  
XX  
XX (U9K) UNIV ROCKEFELLER.  
XX



PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0241826.  
PR 08-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0244674.  
PR 08-NOV-2000; 2000US-024475.  
PR 08-NOV-2000; 2000US-024476.  
PR 08-NOV-2000; 2000US-024477.  
PR 08-NOV-2000; 2000US-024478.  
PR 08-NOV-2000; 2000US-024523.  
PR 08-NOV-2000; 2000US-024524.  
PR 08-NOV-2000; 2000US-024525.  
PR 08-NOV-2000; 2000US-024526.  
PR 08-NOV-2000; 2000US-024527.  
PR 08-NOV-2000; 2000US-024528.  
PR 08-NOV-2000; 2000US-024532.  
PR 08-NOV-2000; 2000US-024609.  
PR 08-NOV-2000; 2000US-024610.  
PR 08-NOV-2000; 2000US-024611.  
PR 08-NOV-2000; 2000US-024613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0251989.  
PR 06-DEC-2000; 2000US-0251719.  
PR 06-DEC-2000; 2000US-0251719.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-502630/55.  
XX P-PSDB: AAM93109.  
XX  
XX Polynucleotides encoding digestive system antigens, useful for  
XX PT diagnosing, treating, preventing and/or prognosing disorders of the  
XX PT digestive system, particularly cancer and cancer metastases -  
XX  
XX Claim 1; SEQ ID NO 1198; 986bp; English.  
XX  
XX The present invention provides the protein and coding sequences of a

CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive  
CC system antigen of the invention.  
XX  
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
  
alignment\_scores:  
Quality: 122.00 Length: 35  
Ratio: 4.067 Gaps: 0  
Percent Similarity: 85.714 Percent Identity: 71.429  
  
alignment block:  
US-08-973-363-8 x AAK88882 ..  
  
Align seq 1/1 to: AAK88882 from: 1 to: 421  
  
1 TleuProaspPaspProaspLysLysProGlnAlaLysGlnLeuGlnH 17  
||||||| :||| :||||||| :|||||||  
110 ATTCTGCCGCTGGAGACAGATMAAAAGCTCAGGGAGAGACTACGAC 159  
  
17 TArgAlaAspTyrLeuLleuLysLeuAsnLysAspLeuAlaArgLysG 34  
||||||| :||||||| :||||||| :||| :|||  
160 CCGAGCGGATTACTTGTGTTGAAGCTGTCAGAAAGGCTCGAGAGAAGG 209  
  
34 LuAla 35  
|||  
210 GGCGT 214  
  
seq\_name: /SIDSI/gcgdata/hold-d-geneseq/geneseg-n-emb1/NA2001A.DAT:AA157603  
seq\_documentation\_block:  
ID AA157603 Standard; cDNA; 421 BP.  
XX  
XX AA157603;  
AC  
XX  
XX 19-OCT-2001 (first entry)  
DT  
XX  
XX Human colorectal cancer antigen cDNA SEQ ID NO: 67.  
DE  
XX  
XX Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200155350-A1.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01350.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0232081.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239335.  
 PR 13-OCT-2000; 2000US-0239337.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0244647.  
 PR 08-NOV-2000; 2000US-0244675.  
 PR 08-NOV-2000; 2000US-0244676.  
 PR 08-NOV-2000; 2000US-0244677.  
 PR 08-NOV-2000; 2000US-0244678.

PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246539.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 06-DEC-2000; 2000US-0251679.  
 PR 06-DEC-2000; 2000US-0251679.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0251990.  
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 PI  
 DR WPI: 2001-45727/49.  
 DR  
 DR P-PSDB; AAM38625.  
 XX  
 XX  
 PT Isolated polypeptide for treating, preventing and/or prognosing  
 PT disorders related to the colon and rectum including colorectal cancers  
 PT and also for testing and detection e.g. diagnosis -  
 XX  
 XX  
 PS Claim 1: SEQ ID NO: 67; 522pp + Sequence Listing; English.  
 CC  
 CC  
 CC The present invention provides the protein and coding sequences of a  
 CC number of colorectal cancer antigens. These are shown in  
 CC AA157647-AA157619 and AAM38569-AA38641. These can be used in the  
 CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.  
 CC The present sequence is a colorectal cancer antigen coding sequence of  
 CC the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcr\_sequences.  
 XX  
 XX  
 SO Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;

alignment\_scores:

Quality: 122.00  
 Ratio: 4.067

Length: 35  
 Gaps: 0

Percent Similarity: 85.714 Percent Identity: 71.429

alignment\_block:  
US-08-973-363-8 x AAI57603 ..

Align seg 1/1 to: AAI57603 from: 1 to: 421

```
1 1lleuProAspAspProaspIysProGlnAlaIysGlnLeuGlnrh 17
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
110 ATTGCGCGGTGGAGACAGATATAAAGCCTCAGGGGAGACACTACAGAC 159
17 rArgAlaAspTYrLeuIleLysLeuAunLysAspLeuAlaArgLysG 34
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
160 CCGACCGGATTAATTCTGTGACAGCTCTCAGAAAGGCTCTGAGAGAAAGG 209
34 LuAla 35
|||
210 GGGCT 214
```

seq\_name: /SIDSL/gcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06443

seq\_documentation\_block:

ID ABL06443 standard; cDNA: 6240 BP.

```
XX AC ABL06443:
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX WO200171042-A2.
XX
XX PN 27-SEP-2001.
XX
XX PD 23-MAR-2001; 2001WO-US09231.
XX
XX PF 23-MAR-2000; 2000US-191637P.
XX
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX DR P-PSDB; ABB62340.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX PS Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;
```

alignment\_scores:

Quality: 88.00 Length: 38  
Ratio: 2.839 Gaps: 2  
Percent Similarity: 81.579 Percent Identity: 55.263

alignment\_block:  
US-08-973-363-8 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

```
1 1lleuProAspAspProaspIysProGlnAlaIysGlnLeuGlnrh 17
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4184 ATTTCCTTAACGAT...ACGGCAAGCCCGAGCCCAACAGCTGCGAGAC 4230
17 rArgAlaAspTYrLeuIleLysLeuAunLysAspLeu...AlaArgL 33
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4231 GCGTCCGAGTACTGCTCTCAGATCATCAGAGAACGTGGAGCTGACCA 4280
33 ysgLuAlaGlnArg 37
|| :|||
4281 ACGGAGACACACGC 4294
```

seq\_name: /SIDSL/gcdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06442

seq\_documentation\_block:

ID ABL06442 standard; cDNA: 9933 BP.

```
XX AC ABL06442:
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX WO200171042-A2.
XX
XX PN 27-SEP-2001.
XX
XX PD 23-MAR-2001; 2001WO-US09231.
XX
XX PF 23-MAR-2000; 2000US-191637P.
XX
XX PR 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX DR P-PSDB; ABB62339.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX PS Claim 1; SEQ ID NO 13808; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX SQ Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;
```

alignment\_scores:                   Quality: 88.00                   Length: 38  
                                   Ratio: 2.839                   Gaps: 2  
                                   Percent Similarity: 81.579                   Percent Identity: 55.263

alignment\_block:  
 US-08-973-363-8 x AB106442 ..

Align seg 1/1 to: AB106442 from: 1 to: 9933

1   lileuProaspapProaspLysylsProglAlaLysGlnLeuGlnh 17  
       |||||                   |||||                   |||||  
 6724   ATTGTGCTTAACGAT...ACGCCACAGCCCCAGCCAGCAGCTGCACGAC 6770

17   rArGAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeu...AlaArgL 33  
       |||||                   |||||                   |||||  
 6771   GCGTGCAGCTACCTGCTCAGATCATCATCAGAGAACGCTGAGCTGACCA 6820

33   ySGlUAAlaArg 37  
       ||                   |||||  
 6821   ACGGAGACGACACCGC 6834

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AA559590

seq\_documentation\_block:  
 ID   AA559590 standard: DNA: 12928 BP.  
 AC   AA559590;  
 XX     
 DT   13-FEB-2002 (first entry)  
 DE     
 XX   Propionibacterium acnes immunogenic protein encoding DNA #85.  
 XX     
 KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW   dermatological; osteopathic; neuroprotectant; ds.  
 OS   Propionibacterium acnes.  
 XX     
 PN   WO200181581-A2.  
 XX     
 PD   01-NOV-2001.  
 XX     
 PE   20-APR-2001; 2001WO-US12865.  
 XX     
 PR   21-APR-2000; 2000US-199047P.  
 PR   02-JUN-2000; 2000US-208841P.  
 PR   07-JUL-2000; 2000US-216747P.  
 XX     
 PA   (CORI-) CORIXA CORP.  
 XX     
 PI   Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI   L'Heuveneure J, Zhang Y, Jen S, Carter D;  
 DR   WPI: 2001-616774/71.  
 XX     
 PT   Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT   vaccinating against and diagnosing infections, especially useful for  
 PT   treating acne vulgaris -  
 XX     
 PS   Claim 1: SEQ ID No 85; 1069PP; English.  
 XX     
 CC   Sequences AA559506-AA559804 represent DNA molecules encoding  
 CC   Propionibacterium acnes immunogenic polypeptides. The proteins and their  
 CC   associated DNA sequences are used in the treatment, prevention and  
 CC   diagnosis of medical conditions caused by P. acnes. The disorders include  
 CC   SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and  
 CC   osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
 CC   in infections of bone, joints and the central nervous system, however it  
 CC   is particularly involved in the inflammatory lesions associated with acne  
 CC   vulgaris. A method for detecting the presence or absence of P. acnes in a

CC   patient comprises contacting a sample with a binding agent that binds to  
 CC   the proteins of the invention and determining the amount of bound protein  
 CC   in the sample. The polypeptides may be used as antigens in the production  
 CC   of antibodies specific for P. acnes proteins. These antibodies can be  
 CC   used to downregulate expression and activity of P. acnes polypeptides and  
 CC   therefore treat P. acnes infections. The antibodies may also be used as  
 CC   diagnostic agents for determining P. acnes presence, for example, by  
 CC   enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
 CC   polypeptides shown in AAU58255-AAU58404 and AAU67601-AAU67602.  
 CC   Note: The sequence data for this patent did not form part of the printed  
 CC   specification, but was obtained in electronic format directly from WIPO  
 CC   at ftp.wipo.int/pub/published\_pot\_sequences.  
 XX     
 SQ   Sequence 12928 BP: 2564 A; 4099 C; 4037 G; 2225 T; 3 other;

alignment\_scores:                   Quality: 61.00                   Length: 40  
                                   Ratio: 2.103                   Gaps: 1  
                                   Percent Similarity: 72.500                   Percent Identity: 32.500

alignment\_block:  
 US-08-973-363-8 x AA559590 ..

Align seg 1/1 to: AA559590 from: 1 to: 12928

3   ProaspapProaspLysylsProglAlaLysGln.....LeuGlnh 17  
       |||||                   |||||                   |||||  
 7295   CCGGACGCTCCATCCGCCACCCCGACCCGCAAGACGCTACTTTCAC 7344

17   rArGAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
       |||||                   |||||                   |||||  
 7345   GTCATCGACGCGTACTCGACGATCTCCGAAAGATGTTCCCTGACCA 7394

34   lUAAlaGlnArgLeuAlaGly 40  
       |||||                   |||||  
 7395   CTGGAGACGCTGTAGCGCGGA 7414

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AA5596071

seq\_documentation\_block:  
 ID   AA5596071 standard: cDNA: 1908 BP.  
 AC   AA5596071;  
 XX     
 DT   26-FEB-2002 (first entry)  
 DE     
 XX   Leishmania antigen 4G2-83 extended DNA.  
 XX     
 KW   Leishmaniasis; Leishmania antigen; immunostimulant; protozoacide; Lb23;  
 KW   interleukin-15; PCR primer; Lbhp83; M15; Lt-1; LbeIF4A; LmSP19; ss;  
 KW   MAPS-1A; LmSP1; LmSP3; LmSP5; LmSP8; LmSP9; LmSP13; LmSP19; ss;  
 KW   LcGSP1; LcGSP3; LcGSP4; LcGSP8; LcGSP10; LcG-34; Lb6-44; 4A5-63; 1B11-39;  
 KW   2A10-37; 4G2-83; 4H6-41; 8G3-100.  
 XX     
 OS   Leishmania major.  
 XX     
 PN   WO200179276-A2.  
 XX     
 PD   25-OCT-2001.  
 XX     
 PE   05-APR-2001; 2001WO-US11254.  
 XX     
 PR   14-APR-2000; 2000US-0551974.  
 PR   05-MAY-2000; 2000US-0565501.  
 PR   14-AUG-2000; 2000US-0639206.  
 XX     
 PA   (CORI-) CORIXA CORP.  
 XX     
 PI   Reed SG, Campos-Neto A, Webb JR, Dillon DC, Skelky YAW, Bhatia A;  
 PI   Coler RM, Probst P;  
 DR   WPI: 2002-061971/08.

DR P-PSDB; AAU71861.  
 XX  
 PS New isolated Leishmania antigens, useful for prevention, treatment and  
 PT diagnosis of leishmaniasis, also related nucleic acids for genetic  
 PT vaccination  
 XX  
 PS Claim 5; Page 190; 193pp; English.  
 XX  
 CC The invention relates to polypeptides comprising an immunogenic part of a  
 CC Leishmania antigen. The leishmania polypeptides and their associated DNA  
 CC sequences, epitopes and fusion proteins are used in the production of  
 CC compositions used for inducing a protective immune response against  
 CC leishmaniasis, for prevention and treatment of the disease. The  
 CC compositions can also be used generally to treat diseases that respond to  
 CC interleukin-15 stimulation. In addition, the products may contain an  
 CC immunostimulant. The sequences represent cDNA molecules encoding  
 CC Leishmania antigens and PCR primers used to amplify DNA of the invention.  
 XX  
 SQ Sequence 1908 BP; 386 A; 531 C; 627 G; 358 T; 6 other;  
 alignment\_scores:  
 Quality: 60.00 Length: 19  
 Ratio: 3.529 Gaps: 0  
 Percent Similarity: 89.474 Percent Identity: 52.632  
 alignment\_block:  
 US-08-973-363-8 x AAS96071 ..  
 Align seg 1/1 to: AAS96071 from: 1 to: 1908  
 2 LeuproaspaspProaspLysLysProGlnAlaLysGlnLeuGlnrHar 18  
 |||||:::|||||::: |||:::|||||::: ||  
 352 CTTCCTGATCAACCTGATCGACCTCCCGGACAGCTGACTTCAGCTCCG 401  
 18 gAlaASP 20  
 |:::|  
 402 AGGTGAC 408  
 seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ30163  
 seq\_documentation\_block:  
 ID AAZ30163 standard; DNA; 34094 BP.  
 XX  
 AC AAZ30163;  
 XX  
 DT 26-JAN-2000 (first entry)  
 XX  
 DE Complete nucleotide sequence of the PAV-3 genome.  
 XX  
 KW PAV-3; defective recombinant PAV vector; live recombinant virus;  
 KW subunit vaccine; nucleic acid immunisation; gene therapy;  
 KW genetic disease; hemophilia; cystic fibrosis; cancer; viral infection;  
 KW acquired immune deficiency syndrome; PAV antigen; porcine pathogen; ds.  
 XX  
 OS Porcine adenovirus Type 3.  
 XX  
 PN WO9953047-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 15-APR-1999; 99WO-US08220.  
 XX  
 PR 15-APR-1998; 98US-0081882.  
 XX  
 PA (UYSA-) UNIV SASKATCHEWAN.  
 XX  
 PI Reddy PS, Tikoo SK, Babiuk LA;  
 XX  
 DR WPI; 1999-620422/53.  
 XX  
 PT New nucleic acids from the genome of porcine adenovirus-3, and derived  
 PT gene therapy vectors, particularly for immunization

XX  
 XX Example 2; Fig 1; 87pp; English.  
 XX  
 CC The present sequence represents the complete nucleotide sequence of the  
 CC genome of porcine adenovirus-3 (PAV-3). The specification also describes  
 CC a defective recombinant PAV vector comprising inverted terminal repeats  
 CC (ITR), packaging sequences and at least one heterologous nucleotide  
 CC sequence (II), but lacking E1 functions. The defective vectors replicate  
 CC inefficiently in cells (other than helper cells) so are unlikely to be  
 CC immunogenic. Deletion of the E1 (and optionally other regions) increases  
 CC the size of heterologous insert that can be packaged. The PAV-3  
 CC polynucleotide sequences are used to produce (recombinant or defective)  
 CC vectors that can express heterologous proteins, e.g. for making live,  
 CC recombinant virus or subunit vaccines, for nucleic acid immunisation or  
 CC for gene therapy (e.g. of genetic diseases such as hemophilia or cystic  
 CC fibrosis, cancer, or viral infections, including acquired immune  
 CC deficiency syndrome), also for in vitro expression of recombinant  
 CC antigens (for antibody production), antisense RNA, ribozymes or  
 CC therapeutic proteins. They are also used diagnostically to detect PAV  
 CC antigens and/or nucleic acid. The vectors may be used in human or  
 CC veterinary medicine, but particularly for expressing protective  
 CC determinants of porcine pathogens. Regulatory regions may be used to  
 CC control expression of heterologous genes. Antibodies raised against PAV-3  
 CC polypeptides can also be used for diagnosis (to detect PAV-specific  
 CC antigen).  
 XX  
 SQ Sequence 34094 BP; 6240 A; 11070 C; 10693 G; 6091 T; 0 other;  
 alignment\_scores:  
 Quality: 58.50 Length: 32  
 Ratio: 2.786 Gaps: 1  
 Percent Similarity: 65.625 Percent Identity: 46.875  
 alignment\_block:  
 US-08-973-363-8 x AAZ30163/rev ..  
 Align seg 1/1 to reverse of: AAZ30163 from: 1 to: 34094  
 5 AspProaspLysLysProGlnAlaLysGlnLeuGlnrHarArgAlaAspTY 21  
 ||| ||| |||||:::|||||::: |||||:::|||||::: |||||  
 7211 GATCAGGAGCGGCTTCGCCACGACGACTGGACACGCGATCAGAGATA 7162  
 21 rLeuIlleLysLeuLeuAsnLysAspLeuAlaArgLysGlnLagln 36  
 | ||||| ||||| |||||:::|||||::: |||||:::|||||::: |||||  
 7161 C.....CTCCTCAACAGACGACCTGTGGAAAGAGAGGCCAG 7125  
 seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:AAZ39679  
 seq\_documentation\_block:  
 ID AAZ39679 standard; DNA; 2911 BP.  
 XX  
 AC AAZ39679;  
 XX  
 DT 02-JUL-1999 (first entry)  
 XX  
 DE Renal cancer associated gene.  
 XX  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9904265-A2.  
 XX  
 PD 28-JAN-1999.  
 XX  
 PF 15-JUL-1998; 98WO-US14679.  
 XX  
 PR 22-JUN-1998; 98US-0102322.  
 PR 17-JUL-1997; 97US-0896164.  
 PR 10-OCT-1997; 97US-0061599.





**THIS PAGE BLANK (USPTO)**



seq. documentation. block: 619 bp mRNA linear EST 18-OCT-2001

LOCUS BBI155356

DEFINITION BBI155356 RIKEN full-length enriched, 16 days neonate thymus Mus musculus cDNA clone A130024L16.3 similar to U10410 Mouse DNA-binding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION BBI155356

VERSION BBI155356.2 GI:16268254

KEYWORDS house mouse.

SOURCE Mus musculus

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 619)

AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyata,T., Muramatsu,M. and Hayashizaki,Y.

TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

JOURNAL Unpublished (2001)

COMMENT On Jun 29, 2000 this sequence version replaced gi:18111286. Contact: Yoshinhide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Saito-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Iawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kirra,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamakaki,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Location/Qualifiers

1. 619

organism="Mus musculus"

db\_xref="taxon:10090"

clone="A130024L16"

clone\_lib="RIKEN full-length enriched, 16 days neonate thymus"

tissue\_type="thymus"

dev\_stage="16 days neonate"

lab\_host="DH10B"

note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGGAGAAGACATCCACAGACTCTTTTGTTCCTTTTTTTTNNV 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to KOT = 10.0 and subtraction to KOT = 185.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGATTCTCGAATTAATAATATCCCCCCCCCCC  
3']. cDNA was cleaved with XhoI and BamHI. Vector: a  
modified pBluescript KS(+) after bulk excision from Lambda

BASE COUNT      194 a    118 c    161 g    146 t  
ORIGIN

alignment\_scores:  
Quality: 187.00 Length: 41  
Ratio: 4.561 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.805

alignment\_block:  
US-08-973-363-8 x BB155356 ..

Align seg 1/1 to: BB155356 from: 1 to: 619

1 lileuPrcAsPaSPPrOasPlYslYsProGlAlALySGlnleuclnTh 17  
|||||  
489 ATTCTTCGCAGATCATCCTGATAAAACCACAAGCAAAACAGTTACAGAC 538  
|||  
17 rrrgalaiaAPTYrleuuilLysleueusnlYsaSplaLaualargLYsg 34  
|||||  
539 CCCTGCAGACTACCTCATCAACTACTTAGCACAGATCTTGCAAAGAG 588  
|||  
34 luAlaclnargleuaLaglyla 41  
|||||  
589 AGCCTCAGAGACTTGTGGTGGC 611  
seq\_name: gb\_estl:AL644594

seg\_documentation\_block:  
LOCUS AL644594 645 bp mRNA linear EST 07-NOV-2001  
DEFINITION AL644594 XGC-egg silurana tropicalis cdna clone lIE1d12 5', mRNA sequence.  
ACCESSION AL644594  
VERSION AL644594.1 GI:16796719  
KEYWORDS EST.  
SOURCE western clawed frog.  
ORGANISM Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Silurana.  
REFERENCE 1 (bases 1 to 645)  
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.  
AUTHORS Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
TITLE Unpublished (2001)  
JOURNAL Contact: Huckle E  
COMMENT Sanger Centre  
Hinnton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: LIE1d12.p1c  
Sequencing primer: PLC  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.

FEATURES  
Source location/Qualifiers  
1..645 /organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="LIE1d12"



seq\_name: gb\_est1:BB834922

seq\_documentation\_block:

LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
Jy9-MC(B) cDNA Mus musculus cDNA clone G930033021 3', mRNA  
sequence.

ACCESSION BB834922  
VERSION BB834922  
KEYWORDS GI:17013165

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 446)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii  
Y., Ito,M., Kawai,T., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse full-length cDNAs (Akimura,T., et al.  
2001)

JOURNAL  
COMMENT Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp.  
URL: http://genome-gsc.riken.go.jp/

Carinci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10  
(11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome-gsc.riken.go.jp) for  
further details.  
e mouse tissues.

FEATURES  
source

1..446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930033021"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jy9-MC(B)"  
BASE COUNT 139 a 99 c 108 g 100 t  
ORIGIN

alignment\_scores:  
Quality: 182.00 Length: 41  
Ratio: 4.439 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 82.927

alignment\_block:

US-08-973-363-8 x BB834922 ..

Align seg 1/1 to: BB834922 from: 1 to: 446

1 l l e l e u p r o a s p a s p p r o a s p l y s l y s p r o g l n a l a l y s i n l e u g l n t h 17  
|||||  
309 a t t c t t c c a g a t g a t c c t g a t a a a a a c c a a g c c a a c a g t t a c a g a c 358  
17 r a r g a l a a s p t y r l e u l l e y s t e u l e u a s n l y a s p l e u a l a r g y s g 34  
|||||  
359 c c g t c a g a c t a c t c c t c a a a c t a c c a g a g a g t t g c c a a a g a g 408  
409 a g g c t c a g a c a c t t t g c g t t c g 431  
34 l u a l a g l n a r g l e u a l a g l y a l a 41  
|||||

seq\_name: gb\_est1:BB830730

seq\_documentation\_block:

LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527  
Jy9-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA  
sequence.

ACCESSION BB830730  
VERSION BB830730  
KEYWORDS GI:17008973

SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 438)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii  
Y., Ito,M., Kawai,T., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse full-length cDNAs (Akimura,T., et al.  
2001)

JOURNAL  
COMMENT Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp.  
URL: http://genome-gsc.riken.go.jp/

Carinci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10  
(11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome-gsc.riken.go.jp) for  
further details.  
e mouse tissues.

```

FEATURES
  source
    Location/Qualifiers
      1..438
        /organism="Mus musculus"
        /db_xref="taxon:10090"
        /clone_id="G330013K04"
        /clone_lib="RIKEN full-length enriched, mammary gland
        RCB-0527 Jy9-MC(B) cDNA"
        /tissue_type="mammary gland"
        /cell_line="RCB-0527 Jy9-MC(B)"
        96 c
        108 g
        96 t
BASE COUNT
  138 a
  96 c
  108 g
  96 t
ORIGIN

alignment_scores:
  Quality: 172.00      Length: 42
  Ratio: 4.195        Gaps: 1
  Percent Similarity: 97.619   Percent Identity: 83.333

alignment_block:
  US-08-973-363-8 x BB830730 ..
  Align seg 1/1 to: BB830730 from: 1 to: 438

1 11leuPProaSPaSPProaSPlySLySPROGINAlaLySGInleuGInth 17
|||||
300 ATTCTTCAGATGATCCGATRAAAAAACCAAGCAAAAGCTTACAGAC 349
|||||
17 TATGAlaSPYrLeuILeLySLeuaAnLySaSPleuAlaLySG 33
|||||
350 CCGTCAGACTACCTCTCAACTGCTTACAGAGATCTTCCAAAAAGA 399
|||||
34 GUAAGlnArGleuAlaGlyAla 41
|||||
400 GAGGCTCAAGACTTGTGTGCG 423

seq_name: gb_est1:AL601246

seq_documentation_block:
  LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001
  DEFINITION DKE2P313J1040_r1_313 (synonym: h1cc2) Homo sapiens cDNA clone
  DKE2P313J1040_5, mRNA sequence.
  ACCESSION AL601246
  VERSION AL601246.1 GI:15164752
  KEYWORDS EST.
  SOURCE
    human.
    ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
  REFERENCE
    1 (bases 1 to 430)
    Bloecker,H., Boecher,M., Brandt,P., Mewes,W., Well,B. and Wiemann
    S.
    EST (Bloecker,H., Boecher,M., Brandt,P., Mewes,W., Well,B. and
    Wiemann,S.)
    JOURNAL Unpublished (1999)
    COMMENT Contact: Bloecker H
    MISC
      Am Klopferspitz 18a D-82152 Martinsried, Germany
      This is the 5' sequence of the clone insert
      Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
      Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
      sequenced by GBR (National Research Centre for Biotechnology Ltd.,
      Braunschweig/Germany) within the cDNA sequencing consortium of the
      German Genome Project.
      No 5' sequence available.
      This clone (DKE2P313J1040) is available at the RZPD in Berlin.
      Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
      Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
  FEATURES
    source
      1..430
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_id="DKE2P313J1040"
        /clone_lib="313 (synonym: h1cc2)"

FEATURES
  source
    /dev_stage="adult"
    /lab_host="DH10B"
    /note="Vector: pTriplex2; Site_1: SfilA; Site_2: SfilB;
    cDNA-collection"
BASE COUNT
  163 a
  81 c
  86 g
  100 t
ORIGIN

alignment_scores:
  Quality: 170.00      Length: 41
  Ratio: 4.359        Gaps: 1
  Percent Similarity: 95.122   Percent Identity: 85.366

alignment_block:
  US-08-973-363-8 x AL601246 ..
  Align seg 1/1 to: AL601246 from: 1 to: 430

1 11leuPProaSPaSPProaSPlySLySPROGINAlaLySGInleuGInth 17
|||||
218 ATTCTTCAGATGATCCGATRAAAAAACCAAGCAAAAGCTTACAGAC 267
|||||
17 TATGAlaSPYrLeuILeLySLeuaAnLySaSPleuAlaLySG 34
|||||
268 CCGTCAGACTACCTCTCAACTGCTTACAGAGATCTTCCAAAAAG 317
|||||
34 GUAAGlnArGleuAlaGlyAla 41
|||||
318 AAGCT.....CTTCTGAGTGG 334

seq_name: gb_est1:AI890775

seq_documentation_block:
  LOCUS AI890775 547 bp mRNA linear EST 07-MAR-2000
  DEFINITION Wm95f11.x1 NCI-CGAP_Ut2 Homo sapiens cDNA clone IMAGE:2443725_3,
  similar to SW:CHD1_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING
  PROTEIN 1; , mRNA sequence.
  ACCESSION AI890775
  VERSION AI890775.1 GI:5595939
  KEYWORDS EST.
  SOURCE
    human.
    ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
  REFERENCE
    1 (bases 1 to 547)
    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
    Unpublished (1997)
    COMMENT Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-remail.nih.gov
    Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
    Emmert-Buck, M.D., Ph.D.
    cDNA Library Preparation: Life Technologies, Inc.
    DNA Sequencing by: Greg Lennon, Ph.D.
    DNA Sequencing by: Washington University Genome Sequencing Center
    Clone distribution: NCI-CGAP clone distribution information can be
    found through the I.W.A.G.B. Consortium/LNL at:
    www.bio.jhu.edu/bihrp/image/image.html
    Insert length: 1924
    Seq primer: -40UP from gibco
    High quality sequence stop: 418.
  FEATURES
    source
      1..547
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_id="IMAGE:2443725"
        /clone_lib="NCI-CGAP_Ut2"
        /tissue_type="moderately-differentiated endometrial
        adenocarcinoma, 3 pooled tumors"
        /lab_host="DH10B"
        /note="Organ: Uterus; Vector: pCW-SPORT6; Site_1: SalI;
        Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT."

```

Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"

BASE COUNT 114 a 118 c 85 g 230 t

ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-8 x AI890775/rev ..

Align seg 1/1 to reverse of: AI890775 from: 1 to: 547

1 IIELEUPROASPAPROASPLYSLSPROGLNALALYSGINLEUGIN 17  
|||||  
378 ATTCTTCAGATGATCCCGATTAATAAACCAACAAACAGTTGCAGAC 329  
17 rARGAASPPTYrLEuIIeLysLEuAuSnYsASpLEuAlAArgLYsg 34  
|||||  
328 CGGTGCAGACTACCTCATCAATTACTTAGTAGAGATCTTGCAAAAAAG 279  
34 IuAAGlInArGLEuAlAGLYaLA 41  
|||||  
278 AAGCT.....CTTCTGTGTCGCG 262

seq\_name: gb\_est1:AU125712

seq\_documentation\_block:  
LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.  
ACCESSION AU125712  
VERSION AU125712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 866)  
Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,  
Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and  
Isogai, T.  
HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,  
Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki  
Y., Sugano, S., Isogai, T.)  
Unpublished (2000)

JOURNAL  
COMMENT Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

FEATURES  
source  
1. 866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_lib="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t 2 others

ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:  
US-08-973-363-8 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

1 IIELEUPROASPAPROASPLYSLSPROGLNALALYSGINLEUGIN 17  
|||||  
450 ATTCTTCAGATGATCCCGATTAATAAACCAACAAACAGTTGCAGAC 499  
17 rARGAASPPTYrLEuIIeLysLEuAuSnYsASpLEuAlAArgLYsg 34  
|||||  
500 CGGTGCAGACTACCTCATCAATTACTTAGTAGAGATCTTGCAAAAAAG 549  
34 IuAAGlInArGLEuAlAGLYaLA 41  
|||||  
550 AAGCT.....CTTCTGTGTCGCG 566

seq\_name: gb\_est2:BE895133

seq\_documentation\_block:  
LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060P1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.  
ACCESSION BE895133  
VERSION BE895133.1 GI:10358221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1028)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs@mail.nih.gov  
Tissue Procurement: ATCC/DCTD/DRP  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMNL at:  
http://image.llnl.gov  
Plate: LLM9753 row: h column: 16  
High quality sequence stop: 488.

FEATURES  
source  
1. 1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_lib="NIH-MGC\_72"  
/clone\_type="melanotic melanoma"  
/tissue="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1; NotI;  
Site: 2; SalI; Cloned unidirectionally. Primer: oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 387 a 205 c 238 g 198 t

ORIGIN

alignment\_scores:  
Quality: 170.00 Length: 41  
Ratio: 4.359 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 85.366

alignment\_block:



US-08-973-363-8 x BE895133

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 lleauproaspaprosplyslsyproginahalsynleunglnphtr 17  
 |||||||  
 131 ATTCTTCACATGATCCGATTAATAAAACCAACGAAACAGTTGCAGAC 180  
 |||||||  
 17 rargalasptryleuilelyslleuansynlsaspleuallarylgc 34  
 |||||||  
 181 CCGTCGACACTACCTATCAATTAATTAAGATCTTGCATAAAG 230  
 |||||||  
 34 luahaglnargleualaglyala 41  
 |||||||  
 231 AAGCT.....CTTCTGCTGG 247

seq\_name: gb\_gss:CNS04DVG

seq\_documentation\_block:

LOCUS CNS04DVG 856 bp DNA linear GSS 21-MAY-2000

DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone 103K08 of library G from Tetradon nigroviridis, genomic survey sequence.

ACCESSION AL286261.1 GI:8024707

VERSION GSS: genome survey sequence.

KEYWORDS Tetradon nigroviridis.

SOURCE Tetradon nigroviridis

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetradon.

REFERENCE 1 (bases 1 to 856)

AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizesmes,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 856)

AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizesmes,C., Mincker,P., Brotlier,P., Quetier,F., Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 856)

AUTHORS Genoscope.

COMMENT Direct Submission

JOURNAL Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases

COMMENT This sequence is a single read and was generated as part of a large scale clone end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetradon>.

FEATURES

source

1..856

/organism="Tetradon nigroviridis"

/db\_xref="taxon:99883"

/clone="103K08"

/note="Genoscope sequence ID: C08G103BF04LP1-end: 17"

BASE COUNT 84 a 308 c 252 t 206 G 6 others

ORIGIN

alignment\_scores:

Quality: 154.00 Length: 39

Ratio: 4.400 Gaps: 0

Percent Similarity: 89.744 Percent Identity: 79.487

Percent Identity: 79.487

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

2 leauproaspaprosplyslsyproginahalsynleunglnphtr 18  
 |||||||  
 674 CTGCGGATATCTCTATTAAGACCTCCAGGCCAAGACTTACAGACG 625  
 |||||||  
 18 galasptryleuilelyslleuansynlsaspleuallarylgc 35  
 |||||||  
 624 AGCCACTACCTATCAATTAATTAAGATCTTGCATAAAG 575  
 |||||||  
 35 laglnargleualagly 40  
 |||||||  
 574 CCCACACACAGSGGSGC 558

seq\_name: gb\_est2:BF239967

seq\_documentation\_block:

LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000

DEFINITION 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:413129 5', mRNA sequence.

ACCESSION BF239967

VERSION BF239967.1 GI:11153890

KEYWORDS EST.

SOURCE human.

ORGANISM Homo.sapiens

REFERENCE 1 (bases 1 to 821)

AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

plate: JLMG103 row: k column: 18

High quality sequence stop: 562.

FEATURES

source

1..821

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:413129"

/clone\_lib="NIH\_MGC\_54"

/tissue\_type="from chronic myelogenous leukemia"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: Bone marrow; Vector: pONR-LIB (Clontech); Site:1: Still (ggcgctcgcc); Site:2: Still (ggcattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCGCGAGCGCGCATG-dT(30)BN-3' (where B = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 345 a 149 c 190 g 137 t

ORIGIN

alignment\_scores:

Quality: 143.00 Length: 36

Ratio: 4.206 Gaps: 1

Percent Similarity: 94.444 Percent Identity: 83.333

Percent Identity: 83.333

Percent Identity: 83.333

Percent Identity: 83.333

Percent Identity: 83.333

Percent Identity: 83.333

Percent Identity: 83.333

Percent Identity: 83.333

Percent Identity: 83.333

Align seg 1/1 to: BF239967 from: 1 to: 821

```

6 ProAspLysIysProGlnAlaIysGlnLeuGlnThrArgAlaAspTyrIle 22
|||||
3 CCGGATTAATAAAACCAACGAAACAGTTCACAGCCGTCACAGACTRACCT 52
|||||
22 uileuLysLeuLeuAsnLysAspLeuAlaArgIysGlnAlaGlnArgLeuA 39
|||||
53 CATCAATATTAATCTAGTAGACATCTTGCAAAAAAGAAAGCT.....CTTT 96
|||||
39 laGlyAla 41
:::|||||
97 CTGGTCGC 104

```

seq\_name: gb\_est1:AW996787

seq\_documentation\_block:

LOCUS AW996787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-403 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW996787  
VERSION AW996787.1 GI:8257021

KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 337)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

AUTHORS

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-qv3-BN0047-230  
200-102-d03&et3=2000-02-23&et4=1)

Seq primer: puc 18 forward  
High quality sequence start: 2  
High quality sequence stop: 337.

Location/Qualifiers

1..337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"

/note="Organ: breast,normal; Vector: puc18; Site:1: Smal;  
Site:2: Smal; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t  
ORIGIN

alignment\_scores: Quality: 133.50 Length: 36  
Ratio: 3.926 Gaps: 2  
Percent Similarity: 94.444 Percent Identity: 83.333

alignment\_block:

us-08-973-363-8 x AW996787/rev ...

Align seg 1/1 to reverse of: AW996787 from: 1 to: 337

```

1 lileuProAspLysIysProGlnAlaIysGlnLeuGlnThrArgAlaAspTyrIle 17
|||||
105 ATTCTTCACATCATGATCCCGATTAATAAAACCAACGAAACAGTTCGACAGAC 56
|||||
17 rArgAlaAspTyrLeuIleLys..LeuLeuAsnLysAspLeuAlaArgIys 33
|||||
55 CCGTGACAGACTATCTCATCAATACTT...AGTAGAGATCTTGCAAAAAA 9
|||||
34 GluAla 35
|||||
8 GAAAGCT 3

```

seq\_name: gb\_est1:AW997058

seq\_documentation\_block:

LOCUS AW997058 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW997058  
VERSION AW997058.1 GI:8257292

KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 686)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.P.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

AUTHORS

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

COMMENT

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-qv3-BN0047-150  
400-152-c03&et3=2000-04-15&et4=1)

Seq primer: puc 18 forward  
High quality sequence start: 19  
High quality sequence stop: 678.

Location/Qualifiers

1..686  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"

/note="Organ: breast,normal; Vector: puc18; Site:1: Smal;  
Site:2: Smal; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 165 a 154 c 126 g 241 t  
ORIGIN

## alignment\_scores:

Quality: 127.00 Length: 35  
Ratio: 3.629 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 82.857

## alignment\_block:

US-08-973-363-8 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

1 lileuProaspapProaspLysLysProGlnAlaLysGlnLeuGlnH 17  
|||||  
127 ATTCTCCAGATGATCCGATTA AAACCAAGCAAAACAGTTCAGAC 79  
|||||  
17 rArgAlaSPtyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
78 CCGTCAGACTACCTCATCAATTAATTAGAGATCTCCAAAAAGAA 29  
34 luAla 35  
|||||  
28 GCTCT 24

seq\_name: gb\_gss:CNS05T9J

seq\_documentation\_block:

LOCUS CNS05T9J 1122 bp DNA linear GSS 26-MAY-2000  
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone  
042M09 of library C from Tetraodon nigroviridis, genomic survey  
sequence.

ACCESSION AL352864  
VERSION AL352864.1 GI:8246657  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetraodon nigroviridis.  
ORGANISM Tetraodon nigroviridis.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 1122)  
AUTHORS Roest-Crolius H., Jallion O., Dasilva C., Fizes C., Fisher C.,  
Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and  
Weissenbach J.

TITLE Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis  
Unpublished

JOURNAL 2 (bases 1 to 1122)

REFERENCE Roest-Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,  
AUTHORS Bernot A., Fizes C., Mincker P., Brotlier P., Quetier F.,  
Saurin W. and Weissenbach J.

Human gene number estimate provided by genome wide analysis using  
Tetraodon nigroviridis DNA sequence  
Unpublished

JOURNAL 3 (bases 1 to 1122)

REFERENCE Genoscope

TITLE Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetraodon>.

Location/Qualifiers

1. 1122

/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone="042M09"

/clone\_lib="C"

/note="Genoscope sequence ID : C0BC042AG05C1-end : T7"

BASE COUNT 303 a 260 c 301 g 246 t 12 others

ORIGIN

## alignment\_scores:

Quality: 124.00

Length: 29

Ratio: 4.276 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 79.310

## alignment\_block:

US-08-973-363-8 x CNS05T9J ..

Align seg 1/1 to: CNS05T9J from: 1 to: 1122

1 lileuProaspapProaspLysLysProGlnAlaLysGlnLeuGlnH 17  
|||||  
835 ATTCTACCAAGATGATCCAGCAAGAGCCCGAGTCCAGCAAGCTACAGGC 884  
|||||  
17 rArgAlaSPtyrLeuIleLysLeuAsnLysAsp 29  
|||||  
885 CAGAGCTGAGTATCTCTCCTCAAGCTGCTGAAAAAGAGC 921





source 1. .6608  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN

alignment\_scores:  
Quality: 197.00 Length: 41  
Ratio: 4.805 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 95.122

alignment\_block:  
US-08-973-363-9 x A58691 ..  
Align seg 1/1 to: A58691 from: 1 to: 6608

1 11eleuproaspaspProaspLyslyspProglinalalysglInleuglnth 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAGAACCCAGCGCAAGCAGCTTACAGAC 4129  
17 rArgalaasPTyrlleuileLysleuenuaslnysAspleuAlaArgLysG 34  
|||||  
4130 CCGTGCAGACTACCTCATTTAAATTACTGAATTAAGACCTTGCAAGAAAG 4179  
34 luValGlnArgLeuThrGlyAla 41  
|||||  
4180 AAGCACAAAGCGCTGCTGCTGCA 4202

seq\_name: gb\_ov:AF004397  
seq\_documentation\_block:  
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the Z chromosome  
protein, variant with hydrophilic domain, (CHD-2) mRNA, complete  
cds.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 6872)  
REFERENCE  
AUTHORS Griffiths, R. and Korn, R.M.  
TITLE A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
2 (bases 1 to 6872)  
REFERENCE  
AUTHORS Griffiths, R. and Korn, R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK

FEATURES  
source  
1. .6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1. .6872  
/gene="CHD-2"  
228..5634  
228..5634  
/gene="CHD-2"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the Z chromosome  
protein"  
/protein\_id="AAC60282.1"  
/db\_xref="GI:2501845"  
/translation="MNGHSDSESVRNSSGSSRDSDGSASGSGSSGSSGSSGSSGSS  
SOGSSSPSESGSGSGSSESSEPSREKKOYQAPPRADGSEPFKSSPSITLAVRSAY  
LKQOQOQKAASDSGSEEDSSSDSDSDSTETKRKRHRDEWQMSGSGSVSGTGS  
DSEAEDEGDRSSCESESDYEPRKNKVRKPPSRKPKSGKSKSTGQKKRQIDSSSEEE

DDDEDYDKRSGRQATVNVSYKEAEETRTDSDLLLEVCGEDVPQTEDEFEETLEKEND  
SRIGRKATGATGATTIYAVEADGDPNAPEFESKELEIOYLKMKGMSHINHTETEET  
LKOONVGMKNLIDNYKKKDDETKRLNASEDEVEYNNCOOELDLDLKHQYOIVERIT  
AHSNOKSAGYPDYCYCKMOGLIPYSECSMEGALIAKFOARIDYEFERNOSKTPFED  
CKVLRKORPRYALAKOPSYIGCHSELRLDYQLNGLWMLAHSKCKSGSLIADBMGIG  
KTIOTISFLNLFHEHQLYGFLLRVPLSTJSMOREIQTWAPQMAVVTIGDITSFN  
MIRHEWMHHPQTKRLKFNILTYVEILLKDKSFLGGLNMAFIVDEAHARKINDSLY  
RTLIDFKSNHRLITGTPLQNSLKELSLHAFIPKFFSWEDEEERHGRGREGYAS  
LHKELEPFLARVKKDYKESLIPKAVEIDLRKMEALOKOYKWTILTYNKALSKSGS  
STSGFLNIMELKCCNHCYLIPKPDNEIPNKEOALOHIRSSGKILIDKILIRL  
ERGNRVILIFSQMYMIDIILAEYLKYRPFPPQROLDGSLKGLERKALDHPNAGEDEDC  
FLSTRAGGLGINLASADTVYIFSDNMPONDQAQARAHRTGKKOVNITRLVTKGS  
VEEDILERAKKKNVLDHLVIQMDTGTQVLTHTGSTPSSSPFKBELSA1LKGAEE  
LFKEPEGEDEQPEQMDIDELIKRAETRENERGPLYVGDDELLSQVNFNSMEDDIE  
LEPERNSRMEIEIPESORRIIEEEROKELEELIYMLPRMNCAKOISFNSGEGRSR  
SRVSGSDSDSITERRKPKKRGPRRTIPREMIKGFSAELIRFTIKSKKRGGS  
AVARDAELVDKSETDLRGLBELVNGCIKALKDSSQGERAGFLGKVKGPTRISY  
QVNAKLVISHHEELAPLHKSTIPSDPEERKRVIPCHRTKAAFDIDWCKEDSNLVI  
YEVYGSWEMIKMPDLSLQKILPDDPKPKVPOKQOTRADYILKLNLDLAKKQ  
RLAGAGNSKRRKTKRNKMKASKIKKEIKSDSPSEKSDSEDEEDENKDELISYK  
HLHKIKTERENEKEPEPDIGIKKEAEERETKEKREKRELKREKEDKELKED  
NKEKREMKVKESTOKEREKVEKYNEMKSEKESKILPILDPVHTATSEVPISL  
ESELHOKTSVCKERRPVAALKOIDREPGLSEBOLEHTKQCIKIGDHTTEL  
KEYTNPDQIKQWRKNLMTFVSKTFEPARKLHLKYKA1KROESQOHNOQNTISNV  
THVTRNDPVERLAKETTHNDDSSRDSYSDRLSQYHDHNDROHGDAYKRSDBKRY  
SAFENGDHRDMWDHYKODSRYSDSKHRKLDHRSNLEGNLSDSGHSPDRSH  
SDRISHSDHRETSEYSHHKSROYRYSDMQMDHRAGSGSGRSPDLDRSPYSGSPUG  
HRSPEHSDHKSPTREHTWSRKT"  
4341..4604  
/gene="CHD-2"  
/note="short insert found in longer variant mRNA of CHD-2"

BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN

alignment\_scores:  
Quality: 197.00 Length: 41  
Ratio: 4.805 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 95.122

alignment\_block:  
US-08-973-363-9 x AF004397 ..  
Align seg 1/1 to: AF004397 from: 1 to: 6872

1 11eleuproaspaspProaspLyslyspProglinalalysglInleuglnth 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAGAACCCAGCGCAAGCAGCTTACAGAC 4129  
17 rArgalaasPTyrlleuileLysleuenuaslnysAspleuAlaArgLysG 34  
|||||  
4130 CCGTGCAGACTACCTCATTTAAATTACTGAATTAAGACCTTGCAAGAAAG 4179  
34 luValGlnArgLeuThrGlyAla 41  
|||||  
4180 AAGCACAAAGCGCTGCTGCTGCA 4202

seq\_name: gb\_pat:A58686  
seq\_documentation\_block:  
LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 5 from Patent WO9639505.  
ACCESSION A58686  
VERSION A58686.1 GI:3714249  
KEYWORDS  
SOURCE  
ORGANISM  
unidentified.  
unclassified.  
REFERENCE  
AUTHORS Griffiths, R. and Korn, R. and Tiwari, B.  
TITLE AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS  
JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;

ISIS INNOVATION (GB)  
 COMMENT Other publication AU 5906996 961224.  
 FEATURES Location/Qualifiers  
 SOURCE 1..153  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 60 a 34 c 31 g 28 t  
 ORIGIN  
 Alignment\_scores:  
 Quality: 191.00 Length: 51  
 Ratio: 4.639 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 80.392  
 Alignment\_block:  
 US-08-973-363-9 x A58686 ..  
 Align seg 1/1 to: A58686 from: 1 to: 153  
 seq\_name: gb\_pat:A58684  
 seq\_documentation\_block:  
 LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
 DEFINITION Sequence 3 from Patent WO9639505.  
 ACCESSION A58684  
 VERSION A58684.1 GI:3714247  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 153)  
 AUTHORS Griffiths R. and Tiwari B.  
 TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
 JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
 COMMENT ISIS INNOVATION (GB)  
 FEATURES Other publication AU 5906996 961224.  
 Location/Qualifiers  
 SOURCE 1..153  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 58 a 40 c 31 g 24 t  
 ORIGIN  
 Alignment\_scores:  
 Quality: 182.00 Length: 51  
 Ratio: 4.439 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 76.471  
 Alignment\_block:  
 US-08-973-363-9 x A58684 ..  
 Align seg 1/1 to: A58684 from: 1 to: 153  
 seq\_name: gb\_pat:A58684  
 seq\_documentation\_block:  
 LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
 DEFINITION Sequence 3 from Patent WO9639505.  
 ACCESSION A58684  
 VERSION A58684.1 GI:3714247  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 153)  
 AUTHORS Griffiths R. and Tiwari B.  
 TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
 JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
 COMMENT ISIS INNOVATION (GB)  
 FEATURES Other publication AU 5906996 961224.  
 Location/Qualifiers  
 SOURCE 1..153  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 58 a 40 c 31 g 24 t  
 ORIGIN

1 ATTTTACCTGATGATCCAGCAAGAACCCAGCAAGCAAGCACTACAGAC 50  
 8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24  
 51 CAAGAAACCCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 100  
 24 yLsLeuLeuAsnLysAspLeuAlaArgLysGlnValGlnArgLeuThrGly 40  
 101 AATTACTGATTAAGACCTTGCAAGAAAGAGACCAAGCAAGCAAGCAAGCA 150  
 41 Ala 41  
 151 GCA 153  
 seq\_name: gb\_pat:A58685  
 seq\_documentation\_block:  
 LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998  
 DEFINITION Sequence 4 from Patent WO9639505.  
 ACCESSION A58685  
 VERSION A58685.1 GI:3714248  
 KEYWORDS  
 SOURCE unidentified.  
 ORGANISM unidentified.  
 REFERENCE 1 (bases 1 to 153)  
 AUTHORS Griffiths R. and Tiwari B.  
 TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
 JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;  
 COMMENT ISIS INNOVATION (GB)  
 FEATURES Other publication AU 5906996 961224.  
 Location/Qualifiers  
 SOURCE 1..153  
 /organism="unidentified"  
 /db\_xref="taxon:32644"  
 BASE COUNT 56 a 36 c 31 g 30 t  
 ORIGIN  
 Alignment\_scores:  
 Quality: 182.00 Length: 51  
 Ratio: 4.439 Gaps: 1  
 Percent Similarity: 80.392 Percent Identity: 76.471  
 Alignment\_block:  
 US-08-973-363-9 x A58685 ..  
 Align seg 1/1 to: A58685 from: 1 to: 153  
 seq\_name: gb\_ro:MUSCHDIX  
 seq\_documentation\_block:  
 LOCUS MUSCHDIX 5349 bp mRNA linear ROD 19-SEP-1996  
 DEFINITION Mouse DNA-Binding protein (CHD-1) mRNA, complete cds.  
 ACCESSION L10410 X66028

VERSION	LI0410.1	GI:455014
KEYWORDS	DNA binding protein.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 5349)	
AUTHORS	Delmas,V., Stokes,D.G. and Perry,R.P.	
TITLE	A mammalian DNA-binding protein that contains a chromodomain and an SNE2/SWI2-like helicase domain	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)	
MEDLINE	93211972	
REFERENCE	2 (bases 1 to 5349)	
AUTHORS	Perry,R.P.	
TITLE	Direct Submission	
JOURNAL	Submitted (08-Apr-1993) Robert Perry, The Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA	
COMMENT	On Feb 16, 1994 this sequence version replaced gi:293322.	
FEATURES	Location/Qualifiers	
source	1..5349	

BASE COUNT	1739 a	1067 c	1319 g	1224 t
ORIGIN				

[illegible]

```
alignment_block:
US-08-973-363-9 x MUSCHDX .
```

```

Align seg 1/1 to: MOSCHD1x from: 1 to: 5349

1 IIELEUPROASPAASPPROASPLYLSYSPROGLNALALYSGINLEUGINTH 17
4026 ATTCCTTCGACATGATCCTGTATATAAAAAACCAACCAAAACAGTTACAGAC 4075
17 IATGALIASPTLYLEULIEYLSLEUTSAUNLYSASPLEULIARGLYSG 34
4076 CCCTGCAGACTACTCCTCACTCAAACTACTTACGACGAGATCTGCACAAAAGAG 4125
34 IIVALGINARGLEUTHINGLYALA 41
4126 AGGCTCAGACACTTTGTGTGGCG 4148

```

```

seq_name: gb_pat:A58683

seq_documentation_block:
LOCUS       A58683                153 bp      DNA
DEFINITION  Sequence 2 from Patent WO96339505.
ACCESSION   A58683
VERSION     A58683.1   GI:3714246
KEYWORDS
SOURCE      unidentified.
ORGANISM    unidentified
             unclassified.
REFERENCE   1 (bases 1 to 153)
            Griffiths, R. and Tiwari, B.
            AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN
            BIRDS
            Patent: WO 9639505-A 2 12-DEC-1996;
            ISIS INNOVATION (GB)
            Other Publication AU 5906996 961224.
JOURNAL
COMMENT     Location/Qualifiers
FEATURES
             1..153
             source

```

alignment_scores:		
Quality:	167.00	Length: 5
Ratio:	4.175	Gaps: 1
Percent Similarity:	78.431	Percent Identity: 68.622

```
alignment_block:
US-08-973-363-9 x A58683 ..
```

[illegible]

seq_name: gb_pr:AF006513	
seq_documentation_block:	
LOCUS AF006513	5947 bp mRNA
DEFINITION Homo sapiens CHD1 mRNA, complete cds.	linear
ACCESSION AF006513	
VERSION AF006513.1	GI:2645428
	PRI 27-NOV-1997



[illegible]

```

1 1LELEUTROASPAAPSPROASPLYSLSYSGRGLALALYSGLINLEUGLNTH 17
  |||||||
4025 ATCTCTCCGATGATCCCGCTAAAAAACCAACGAACAAAACGTTGCAGAC 4074

17 FATGALAAAPYIRLEUULEYSLLEULSLEULSLEULSASPLLEUULARYUSG 34
  |||||||
4075 CCGTGCAGACTRACCTCATCAATATTCCTACTAGAACGTTCTTCACAAAAAAG 4124

34 luValGlaRygleuThrGlyAla 41
  ||:::|||||
4125 AAGCT.....CTTTCGTGCTGGC 4141

seq_name: gb_pr:AC092372

seq_documentation_block:
LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-200
DEFINITION Homo sapiens chromosome 5 clone RP11-98M12, complete sequence.
ACCESSION AC092372
VERSION AC092372.3 GI:17402768
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 101220)
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL TITLE Direct Submission
REFERENCE 2 (bases 1 to 101220)
DOE Joint Genome Institute.
AUTHORS TITLE Direct Submission
JOURNAL TITLE Direct Submission
REFERENCE 3 (bases 1 to 101220)
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS TITLE Direct Submission
JOURNAL TITLE Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 7, 2001 this sequence version replaced gi:15290448.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 146,7kb). It is clipped at the overlap with AC012624.
The number of bases overlapped is 90404.
FEATURES
source
1..101220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-98M12"
BASE COUNT 34122 a 18862 c 17827 g 30409 t
ORIGIN

alignment_scores:
Quality: 166.00 Length: 41
Ratio: 4.256 Gaps: 1
Percent Similarity: 95.122 Percent Identity: 82.927

alignment_block:
05-08-975-363-9 x AC092372/rev ..

Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220

1 1LELEUTROASPAAPSPROASPLYSLSYSGRGLALALYSGLINLEUGLNTH 17
  |||||||
26932 ATCTCTCCGATGATCCCGCTAAAAAACCAACGAACAAAACGTTGCAGAC 26883

```

```

17  TARGALASPRTYrleuilelysleuAasnlyAspLeuAlaArglySG 34
|||||
26882 CCGGACGAGTACTCATCAATTAATTAGTAGAGATCTTGCAAAAAAAG 26833
34  luvAlGlnArgleuThrglyAla 41
|||||
26832 AAGCT.....CTTCTGCTGCG 26816
seq_name: gb_pr:AC012624

seq_documentation_block:
LOCUS      AC012624      134365 bp      DNA      linear      PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
ACCESSION  AC012624
VERSION    AC012624.6  GI:14993679
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE  1 (bases 1 to 134365)
            DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS   Direct Submission
TITLE     Unpublished
JOURNAL   2 (bases 1 to 134365)
REFERENCE  DOE Joint Genome Institute.
AUTHORS   Direct Submission
TITLE     Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
JOURNAL   Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134365)
REFERENCE  DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS   Direct Submission
TITLE     Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL   Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 134365)
REFERENCE  DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS   Direct Submission
TITLE     Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL   Drive, Walnut Creek, CA 94598, USA
COMMENT   On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES   Location/Qualifiers
            source          1..134365
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN

```

```

alignment_scores:
    Quality: 166.00      Length: 41
    Ratio: 4.256         Gaps: 1
    Percent Similarity: 95.122  Percent Identity: 82.927

alignment_block:
US-08-973-363-9 x AC012624 ..
Align seg 1/1 to: AC012624 from: 1 to: 134365
1  TleuPProAspPProAspPlySLysProGlnAlaLysGlnleuGlnth 17
|||||
118247 ATTCTTCACATGATGCCGATTAATAAACCAACAAACAGTTGCAGAC 118296
17  TARGALASPRTYrleuilelysleuAasnlyAspLeuAlaArglySG 34
|||||
118297 CCGGACGAGTACTCATCAATTAATTAGTAGAGATCTTGCAAAAAAAG 118346
34  luvAlGlnArgleuThrglyAla 41
|||||
118347 AAGCT.....CTTCTGCTGCG 118363
seq_name: gb_hhg:AC021449

```

```

seq_documentation_block:
LOCUS      AC021449      143079 bp      DNA      linear      HTG 10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION  AC021449
VERSION    AC021449.3  GI:10047806
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE  1 (bases 1 to 143079)
            Birren, B., Linton, L., Nusbaum, C. and Lander, E.
AUTHORS   Homo sapiens, clone RP11-58M12
TITLE     Unpublished
JOURNAL   2 (bases 1 to 143079)
REFERENCE  Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
            Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,
            Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
            Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
            Dearrellano, K., Dewar, K., Domino, M., Doyle, M., Fenesstor, J.,
            Ferrelita, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
            Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
            Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
            Landers, T., Lehoczeky, J., Levine, R., Liu, C., Liu, G., Locke, K.,
            Macdonald, P., Marcus, N., McEwan, P., McGurk, A., McKernan, K.,
            McPheeters, R., Meldrum, J., Menais, L., Morrow, J., Naylor, J.,
            Norman, C. H., O'Connor, T., O'Donnell, P., O'Leary, J. M., Peterson, K.,
            Pierre, N., Pisanu, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
            Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
            Stojanovic, N., Subramanian, A., Talamas, J., Testafaye, S., Theodore, J.,
            Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
            Zimmer, A. and Zody, M.
            Direct Submission
            Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Sep 10, 2000 this sequence version replaced gi:1407963.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/MW/repeatmasker.html
            ----- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence.submissions@genome.wi.mit.edu
            ----- Project Information
            Center project name: L5154
            Center clone name: 58_M_12
            ----- Summary Statistics
            Sequencing vector: M13; M7815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 134743 bases at least Q40
            Consensus quality: 139227 bases at least Q30
            Consensus quality: 140814 bases at least Q20
            Insert size: 14400; agarose-fp
            Insert size: 142179; sum-of-contigs
            Quality coverage: 4.6 in Q20 bases; agarose-fp
            Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            -----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 38820: contig of 38820 bp in length
            * 38821 38920: gap of 100 bp
            * 38921 40411: contig of 1491 bp in length
            * 40412 40511: gap of 100 bp

```



Ratio: 4.256 Gaps: 1  
Percent Similarity: 95.122 Percent Identity: 82.927  
alignment block:  
US-08-973-363-9 x AC008531 ..  
Align seq 1/1 to: AC008531 from: 1 to: 145659  
1 lleleuProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17  
46770 ATTCTTCAGATGATCCGATAAAAACACACAAACAGCTGCAGAC 46819  
17 rArgAlaAspTyrLeuIleuLysLeuAaLnLysAspLeuAlaArgLysG 34  
46820 CCGTGCAGACTACCTCATCAAAATTACTAGAGAGATCTTCAAAAAAG 46869  
34 luValGlnArgLeuThrGlyAla 41  
46870 AAGCT.....CTTCTGTGTCG 46886  
seq\_name: gb\_htg:AC091946  
seq\_documentation block:  
LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
PROCESS \*\*\*, 33 unordered pieces.  
ACCESSION AC091946  
VERSION AC091946.1 GI:143333882  
KEYWORDS HTG: HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Unpublished  
Sequencing of Human Chromosome 5  
2 (bases 1 to 193446)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 544799  
Center clone name: RPCI-11\_36012  
-----  
Summary Statistics  
Consensus quality: 157767 bases at least Q40  
Consensus quality: 180259 bases at least Q30  
Consensus quality: 184175 bases at least Q20  
Estimated insert size: 204590; agarose-fp estimation  
Estimated insert size: 190246; sum-of-contigs estimation  
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation  
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1358: contig of 1358 bp in length  
\* 1359 1458: gap of unknown length  
\* 1459 2667: contig of 1209 bp in length  
\* 2668 2767: gap of unknown length  
\* 2768 3823: contig of 1056 bp in length  
\* 3823 3923: gap of unknown length  
\* 3924

5920: contig of 1997 bp in length  
\* 5921 6020: gap of unknown length  
\* 6021 7494: contig of 1474 bp in length  
\* 7495 7594: gap of unknown length  
\* 7595 9096: contig of 1502 bp in length  
\* 9097 9196: gap of unknown length  
\* 9197 10260: contig of 1064 bp in length  
\* 10261 10360: gap of unknown length  
\* 10361 12560: contig of 2100 bp in length  
\* 12561 14611: gap of unknown length  
\* 14612 14712: gap of unknown length  
\* 14712 16381: contig of 1670 bp in length  
\* 16382 16481: gap of unknown length  
\* 16482 17968: contig of 1487 bp in length  
\* 17969 18068: gap of unknown length  
\* 18069 20434: contig of 2366 bp in length  
\* 20435 20534: gap of unknown length  
\* 20535 23515: contig of 2981 bp in length  
\* 23516 23615: gap of unknown length  
\* 23616 27563: contig of 3948 bp in length  
\* 27564 30987: gap of unknown length  
\* 30988 31087: gap of unknown length  
\* 31088 36065: contig of 4978 bp in length  
\* 36066 36165: gap of unknown length  
\* 36166 40978: contig of 4813 bp in length  
\* 40979 41078: gap of unknown length  
\* 41079 45663: contig of 4585 bp in length  
\* 45664 45763: gap of unknown length  
\* 45764 51745: contig of 5982 bp in length  
\* 51746 51845: gap of unknown length  
\* 51846 57359: contig of 5514 bp in length  
\* 57360 57459: gap of unknown length  
\* 57460 67881: contig of 10422 bp in length  
\* 67882 67981: gap of unknown length  
\* 67982 74132: contig of 6151 bp in length  
\* 74133 74232: gap of unknown length  
\* 74233 79795: contig of 5563 bp in length  
\* 79796 79895: gap of unknown length  
\* 79896 87511: contig of 7616 bp in length  
\* 87512 87611: gap of unknown length  
\* 87612 92791: contig of 5160 bp in length  
\* 92792 92891: gap of unknown length  
\* 92892 102794: contig of 9903 bp in length  
\* 102795 102894: gap of unknown length  
\* 102895 110866: contig of 7972 bp in length  
\* 110867 110966: gap of unknown length  
\* 110967 117571: contig of 6605 bp in length  
\* 117572 117671: gap of unknown length  
\* 117672 123738: contig of 6067 bp in length  
\* 123739 130583: gap of unknown length  
\* 123839 130583: contig of 6745 bp in length  
\* 130584 130683: gap of unknown length  
\* 130684 141544: contig of 10861 bp in length  
\* 141545 141644: gap of unknown length  
\* 141645 169109: contig of 27465 bp in length  
\* 169110 169209: gap of unknown length  
\* 169210 193446: contig of 24237 bp in length.  
Location/Qualifiers  
1.193446  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-36012"  
/clone\_11b="RPCI human BAC library 11"

BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others  
ORIGIN

alignment\_scores:  
Quality: 166.00 Length: 41  
Ratio: 4.256 Gaps: 1

Percent Similarity: 95.122 Percent Identity: 82.927

alignment\_block:

US-08-973-363-9 x AC091946

Align seg 1/1 to: AC091946 from: 1 to: 193446

1 HLEuPtoAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
87658 ATTCTTCAGATGATCCGATATAAAACCAAGCAAAAGATTGCGAGC 87707

17 rArGAlaAspTyrLeuLleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
87708 CCGTGCAGACTACCTCATCAATTAATTACTAGAGATCTTGCACAAAAAAG 87757

34 luvAlGlnArgLeuThrGlyAla 41  
|||||  
87758 AAGCT.....CTTCTGGTGCG 87774

seq\_name: gb\_pr:AC026778

seq\_documentation\_block:

LOCUS AC026778 195433 bp DNA linear PRI 01-JUN-2001

DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.

ACCESSION AC026778

VERSION AC026778.4 GI:14277282

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE 1 (bases 1 to 195433)

JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.

REFERENCE Direct Submission

AUTHORS 2 (bases 1 to 195433)

TITLE DOE Joint Genome Institute.

JOURNAL Direct Submission

REFERENCE Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint

AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

TITLE 3 (bases 1 to 195433)

JOURNAL DOE Joint Genome Institute and Stanford Human Genome Center.

REFERENCE Direct Submission

AUTHORS Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell

TITLE Drive, Walnut Creek, CA 94598, USA

JOURNAL On Jun 1, 2001 this sequence version replaced gi:13677045.

REFERENCE Draft Sequence Produced by DOE Joint Genome Institute

AUTHORS www.jgi.doe.gov

TITLE Finishing Completed at Stanford Human Genome Center

JOURNAL www.shgc.stanford.edu

REFERENCE Quality: Phrap Quality >=40 99.9% of Sequence;

AUTHORS Estimated Total Number of Errors is 0.2.

TITLE STS Content:

JOURNAL WI-13675 G23101

REFERENCE SHGC-58345 G38487

COMMENT SHGC-103595 G57841.

## FEATURES

source

1. 195433

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="CTC-428111"

BASE COUNT 62762 a 37302 c 37040 g 58329 t

ORIGIN

alignment\_scores:

Quality: 165.00

Ratio: 4.256

Percent Similarity: 95.122

Percent Identity: 82.927

alignment\_block:

US-08-973-363-9 x AC026778/rev

Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433

1 HLEuPtoAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
29732 ATTCTTCAGATGATCCGATATAAAACCAAGCAAAAGATTGCGAGC 29683

17 rArGAlaAspTyrLeuLleLysLeuLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
29682 CCGTGCAGACTACCTCATCAATTAATTACTAGAGATCTTGCACAAAAAAG 29633

34 luvAlGlnArgLeuThrGlyAla 41  
|||||  
29632 AAGCT.....CTTCTGGTGCG 29616

---



34 luValGlnArgLeuThrGlyAla 41  
 1238 AACGACAGAGACTTGCTGTGCA 1260

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751

seq\_documentation\_block:

ID AA142751 standard; cDNA; 6608 BP.

AA142751;

12-MAR-1997 (first entry)

Chicken CHD-1A gene.

Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

Gallus sp.

Key Location/Qualifiers

FT CDS 228..5390

FT /\*tag= a

W09639505-A1.

12-DEC-1996.

05-JUN-1996; 96WO-GB01341.

06-JUN-1995; 95GB-0011439.

(ISIS-) ISIS INNOVATION LTD.

Griffiths R, Tiwari B;

WPI; 1997-043127/04.

Avian chromodomain-helicase-DNA binding genes determine sex in

birds - used for sex determ. and to control sex of progeny

Claim 1; Fig 5; 76pp; English.

The chicken CHD-W gene (AA142754) acting alone or in conjunction with

the closely related CHD-1A gene (AA142751) is suggested to initiate

female development in birds. The sequence of CHD-1A was deduced

from 3 clones isolated from a stage 10-12 chicken embryo cDNA

library using a great tit CHD-W sequence (see also AA142755) as probe.

The CHD-1A (A - Avian) gene shows close identity to the mouse CHD-1

gene (see also AA142756-57). It is located on an autosome or Z

chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-

specific signal on hybridisation to genomic DNA of a non-ratite

bird and can be used for sex determin. of a bird. CHD-1A nucleic

acids can also be used to control the sex of the progeny of a bird.

Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores: Quality: 197.00 Length: 41

Ratio: 4.805 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 95.122

alignment\_block: US-08-973-363-9 x AA142751 ..

Align seg 1/1 to: AA142751 from: 1 to: 6608

1 lleuProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnrh 17  
 4080 ATTTTACCTGATGATCCAGACAAAGCCAGGCAAGCAGCTACAGAC 4129  
 17 rArgAlaAspTyrIleuIleuLysLeuLeuAsnLysAspLeuAlaArgLysG 34

|||||  
 4130 CCGTGCAGACTACCTCAATTAATTACTGAATAAGACTTGCAAGAAAGC 4179  
 34 luValGlnArgLeuThrGlyAla 41  
 4180 AACGACAAAGGCTTGCTGTGCA 4202

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759

seq\_documentation\_block:

ID AA142759 standard; DNA; 153 BP.

AA142759;

12-MAR-1997 (first entry)

Great tit CHD-W gene fragment.

Bird; sex determination; chromodomain-Helicase-DNA binding 1;

CHD-1A; CHD-W; W chromosome; ss.

Parus major.

Key Location/Qualifiers

FT misc\_difference 52..81

/\*tag= a "bases 52-81 are a repeat of bases 22-51 and are ignored in the translated amino acid sequence given in Fig 3"

W09639505-A1.

12-DEC-1996.

05-JUN-1996; 96WO-GB01341.

06-JUN-1995; 95GB-0011439.

(ISIS-) ISIS INNOVATION LTD.

Griffiths R, Tiwari B;

WPI; 1997-043127/04.

Avian chromodomain-helicase-DNA binding genes determine sex in

birds - used for sex determ. and to control sex of progeny

Claim 8; Fig 3; 76pp; English.

Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology

to portions of the chicken CHD-1A (A - Avian) gene (AA142757),

CC chicken CHD-W (W refers to the W chromosome) gene (AA142758) and

CC and the great tit CHD-W gene (AA142759). Translated amino acid

sequences of this region are provided in AA142754-55. The CHD-1A

(see also AA142751) and CHD-W (see also AA142754-55) genes determine

sex in birds and can be used to identify the sex of an embryo,

foetus etc. and to manipulate the sex of progeny.

Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

alignment\_scores: Quality: 191.00 Length: 51

Ratio: 4.659 Gaps: 1

Percent Similarity: 80.392 Percent Identity: 80.392

alignment\_block: US-08-973-363-9 x AA142759 ..

Align seg 1/1 to: AA142759 from: 1 to: 153

1 lleuProaspaspProasp..... 7



```

|||||
1 ATTTACTGATGATCCAGATAGAGAAACACAGGCAAGCACTTGCAGAC 50
8 .LysLeuProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlel 24
51 CAAGAAACCCAGGCAAGCAAGCTTGCAGACCCGTCGACATTCCTCATTA 100
24 yslLeuLeuAsnLysAspLeuAlaArgLysGlnValGlnArgLeuThrGly 40
101 AATTACTGATATTAAGACCTTGCAAGAAAGAGAGCTGCAAGACCTTCTGT 150
41 Ala 41
151 GCA 153

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142757
seq_documentation_block:
ID AA142757 standard; DNA: 153 BP.
AC AA142757;
XX
XX 12-MAR-1997 (first entry)
DE Chick CHD-1A gene fragment.
XX
XX Bird: sex determination: chromodomain-helicase-DNA binding 1;
KM CHD-1A: CHD-W; W chromosome: ss.
XX
XX Gallus sp.
OS
XX
XX Key Location/Qualifiers
FH misc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX
XX P-PSDB: AAW08147.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757).
XX chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)
XX and the great tit CHD-W gene (see also AA142759). Translated amino
XX acid sequences of this region are provided in AAW08146-49. The
XX CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes
XX determine sex in birds and can be used to identify the sex of an
XX embryo, foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment_scores:
Quality: 182.00 Length: 51
Ratio: 4.439 Gaps: 1

Percent Similarity: 80.392 Percent Identity: 76.471
Alignment_block:
US-08-973-363-9 x AA142757
Align seq 1/1 to: AA142757 from: 1 to: 153

1 TleLeuProAspAspProAsp..... 7
1 ATTTACTGATGATCCAGATCCAGAAACCCAGCAAGCACTTGCAGAC 50
8 .LysLeuProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlel 24
51 CAAGAAACCCAGGCAAGCAAGCTTGCAGACCCGTCGACATTCCTCATTA 100
24 yslLeuLeuAsnLysAspLeuAlaArgLysGlnValGlnArgLeuThrGly 40
101 AATTACTGATATTAAGACCTTGCAAGAAAGAGAGCTGCAAGACCTTCTGT 150
41 Ala 41
151 GCA 153

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758
seq_documentation_block:
ID AA142758 standard; DNA: 153 BP.
AC AA142758;
XX
XX 12-MAR-1997 (first entry)
DE Chick CHD-W gene fragment.
XX
XX Bird: sex determination: chromodomain-helicase-DNA binding 1;
KM CHD-1A: CHD-W; W chromosome: ss.
XX
XX Gallus sp.
OS
XX
XX Key Location/Qualifiers
FH misc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX
XX P-PSDB: AAW08148.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757).
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAW08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,

```

```

CC      foetus etc. and to manipulate the sex of progeny.
XX
S0      Sequence 153 BP: 56 A; 36 C; 31 G; 30 T; 0 other;

alignment_scores:
      Quality: 182.00      Length: 51
      Ratio: 4.439      Gaps: 1
      Percent Similarity: 80.392      Percent Identity: 76.471

alignment_block:
US-08-973-363-9 x AAT42758 ..

Align seg 1/1 to: AAT42758 from: 1 to: 153

      1 1lelePProspAspProksp..... 7
      1 ATTTTTRCCTGATGATCCAGATTAAGAAACCCGAGCTTAAGCACTTACAGAC 50
      8 .LysIysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleL 24
      51 CAAGAAACCCCGAGCGCTTAAGCACTTACAGACCCGTCGAGATTACGCATTA 100
      24 yslenuleuAnslYsAspleuAlaArgLysGlnValAlGlnArgLeuThrGly 40
      101 AATTACTGTGATTAAGACCTTGCAGAAAGAAAGAGACAGAGACTTCTGCTG 150
      41 Ala 41
      |||
      151 GCA 153

seq_name: /SIS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAT42756
seq_documentation_block:
ID AAT42756 standard; DNA; 153 BP.
XX
AC AAT42756;
XX
DT 12-MAR-1997 (first entry)
XX
DE Mouse CHD-1 gene (bases 3855-9777).
XX
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KM CHD-1; CHD-W; W chromosome; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81 /*tag- a
FT FT /note= "bases 52-81 are a repeat of bases 22-51
FT FT and are ignored in the translated amino
FT FT acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96WO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
XX
XX WPI; 1997-043127/04.
XX
XX P-PSDB; AAM08146.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX

```

```

XX Bases 3855-3977 (AAAT42756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (see also
CC and AAAT42757), chicken CHD-W (W refers to the W chromosome) gene
CC (see also AAAT42758) and the great tit CHD-W gene (see also AAAT42759).
CC Translated amino acid sequences of this region are provided in
CC AA08146-49. The CHD-1A (see also AAAT42751) and CHD-W (see also
CC AAAT42754-55) genes determine sex in birds and can be used to identify
CC the sex of an embryo, foetus etc. and to manipulate the sex of
CC progeny.
XX
XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;
XX
XX
XX alignment_scores:
XX      Quality: 167.00      Length: 51
XX      Ratio: 4.175      Gaps: 1
XX Percent Similarity: 78.431      Percent Identity: 68.627
XX
XX alignment_block:
XX US-08-973-363-9 x AAAT42756 ..
XX
XX Align seg 1/1 to: AAAT42756 from: 1 to: 153
XX
XX      1 lEleuProAspAspProAsp.....7
XX      |||||||
XX      1 ATTCTTCAGATGATCCTGATAAAACACCAAGCAAAACACTTACAGAC 50
XX
XX      8 .LysLysProGlnAlaLysGlnLeuGlnThrArgAlaSprYrLeuIleL 24
XX      |||||||
XX      51 CAAAAAACCAACCAAGCAAAACAGTTACAGACCGCTCCAGACTACCTCATCA 100
XX
XX      24 yslEuleuAaNslySaspleuAlaArgLysGlnValGlnArgLeuThrGly 40
XX      |||||||
XX      101 AACTACTTAGCAGAGATCTTGCCAAAAAGAGAGGCTCAGAGACTTTGTGGT 150
XX
XX      41 Ala 41
XX      |||
XX      151 GCG 153
XX
XX seq_name: /SIBS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV59280
XX
XX seq_documentation_block:
XX ID AAV59280 standard; cDNA: 1311 BP.
XX
XX AAV59280;
XX AC
XX XX
XX 14-DEC-1998 (first entry)
XX DT
XX XX
XX DE Altered telomere repeat binding factor 1 gene.
XX
XX XX
XX ds: human; telomere repeat binding factor; A-TRF; dimerisation domain;
XX telomere; aging; ataxia telangiectasia; Down's syndrome; tumour; viral
XX RW
XX XX
XX Homo sapiens.
XX OS
XX XX
XX Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT CDS 1..1311
XX TT /*tag= a
XX TT /product= "A-TRF"
XX
XX PN W09836066-A1.
XX
XX XX
XX 20-AUG-1998.
XX PD
XX
XX PF 13-FEB-1998; 98MO-US02765.
XX
XX XX
XX 04-FEB-1998; 98US-0018628.
XX PR
XX 13-FEB-1997; 97US-0800264.
XX
XX XX
XX (UVRQ ) UNIV ROCKEFELLER.
XX

```

PI Bianchi A, De Lange T, Van Steensel B;  
 XX MPI: 1998-480769/41.  
 DR P-ESDB: AAV59280.  
 XX  
 PT Nucleic acid encoding altered telomere repeat binding protein and  
 related vectors - transformants, hetero-dimers and antibodies, used  
 PT to inhibit shortening of telomerases caused by ageing or disease,  
 PT also used to extend life of cells in culture  
 XX  
 PS Claim 14: Page 110-111; 163pp; English.  
 XX  
 CC The altered vertebrate telomere repeat binding protein (A-TRF) has a  
 CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
 CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
 CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
 CC inhibit shortening of telomeres associated with ageing (for cosmetic  
 CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
 CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
 CC tumours and viral (including human immunodeficiency virus) infection.  
 CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
 CC for expression of recombinant proteins or where intended for subsequent  
 CC transplant or for testing, eliminating the need for transformation.  
 CC  
 SO Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other:  
 alignment\_scores:  
 Quality: 129.00 Length: 25  
 Ratio: 5.160 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-9 x AAV59280 ..  
 Align seg 1/1 to: AAV59280 from: 1 to: 1311  
 1 TleleuproaspaspProaspPylsPylsPglAlaLysGlnLeuGlnth 17  
 |||||  
 1237 ATTCTTCAGATGATCTCATTAACCAAGCAAGCAAGCAAGTACAGAC 1286  
 |||||  
 17 rArGaLaSPtYrLeuIleLysIleu 25  
 |||||  
 1287 CCGTCACACATCTCATCAACTTA 1311  
 seq\_name: /SIBS1/gcgcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK88882  
 seq\_documentation\_block:  
 ID AAK88882 standard; cDNA; 421 BP.  
 XX  
 AC AAK88882;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human digestive system antigen coding sequence SEQ ID NO: 1198.  
 XX  
 KM Human digestive system antigen; gene therapy; cancer; appendicitis;  
 KM ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
 KM digestive system disorder; Meckel's diverticulum; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155314-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PE 17-JAN-2001; 2001WO-US01324.  
 XX  
 PR 31-JAN-2000; 2000US-0179055.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198122.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0205467.  
 PR 28-JUN-2000; 2000US-0214866.  
 PR 30-JUN-2000; 2000US-0215133.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0228287.  
 PR 01-SEP-2000; 2000US-0228343.  
 PR 01-SEP-2000; 2000US-0228344.  
 PR 01-SEP-2000; 2000US-0228345.  
 PR 05-SEP-2000; 2000US-0228509.  
 PR 05-SEP-2000; 2000US-0228513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 12-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235488.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235835.  
 PR 28-SEP-2000; 2000US-0236327.  
 PR 28-SEP-2000; 2000US-0236367.  
 PR 28-SEP-2000; 2000US-0236368.  
 PR 28-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.

PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0249960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241787
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249300
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250391
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251988
PR	05-DEC-2000	2000US-0256719
PR	06-DEC-2000	2000US-0251479
PR	08-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251868
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251989
PR	08-DEC-2000	2000US-0251990
PR	11-DEC-2000	2000US-0254097
PR	05-JAN-2001	2001US-0259678
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI: 2001-502630/55.	
DR	P-PSDB: AAM93109.	
XX		
PT	Polynucleotides encoding digestive system antigens, useful for	
PT	diagnosing, treating, preventing and/or prognosing disorders of the	
PT	digestive system, particularly cancer and cancer metastases -	
XX		
PS	Claim 1; SEQ ID NO 1198; 966bp; English.	
CC	The present invention provides the protein and coding sequences of a	

CC	number f human digestive system antigens. These can be used in the
CC	diagnosis, treatment and prevention of digestive system disorders,
CC	including cancer; Meckel's diverticulum, bacterial or parasitic
CC	infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC	ulcerative colitis. The present sequence is a cDNA encoding a digestive
CC	system antigen of the invention.
XX	
SQ	Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other:
	alignment_scores:
	Quality: 120.00 Length: 33
	Ratio: 4.138 Gaps: 0
	Percent Similarity: 87.879 Percent Identity: 72.727
	alignment_block:
	US-08-973-363-9 x AAK88882 ..
	Align seg 1/1 to: AAK88882 from: 1 to: 421
	1 HlelepProaspasproasplyslsySprogHAlALysSlmLeugnrh 17
	:::          :::
	110 ATPTCTCCGGTGGACAGATAAAGCCCTCAGGGAAGCACTAACAGAC 159
	17 rArgAlAsPtyrLeuIllelslsLeuEuAnslySAsPlleuAlArGLy 33
	:::     :::     :::
	160 CCAGAGCGGATTACTTGTTGAGCTGCAGAAAGGCTGTGCAGAAGAG 208
	seq_name: /SIDS1/cgdata/hold-geneseq/gene-seq-n-emb1/NA2001A.DAT:AAI57603
	seq_documentation_block:
ID	AAI57603 standard; CDNA: 421 BP.
XX	
AC	AAI57603;
XX	
DT	19-OCT-2001 (first entry)
DE	
XX	Human colorectal cancer antigen cDNA SEQ ID NO: 67.
XX	
KW	Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO20015350-A1.
XX	
PD	02-AUG-2001.
XX	
PF	17-JAN-2001; 2001WO-US01350.
XX	
XX	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205513.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215133.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.











alignment\_scores:

Quality:	60.50	Length:	37
Ratio:	2.161	Caps:	1
Percent Similarity:	75.676	Percent Identity:	29.730

alignment\_block:

US-08-973-363-9 x ABI99578

Align seg 1/1 to: ABI99578 from: 1 to: 826

```

2 LeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnThrAf 18
|||||.....|.....|.....|.....|.....|.....|.....|
278 CTTCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 327
18 gAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaargLysGlnV 35
: : : : : | | | | | : : : : : | | | | | : : : : : |
328 ATTTGAGAACCTCTGCAAAATTATG...AAGGATATTGTTGGAGAGAGAG 374
35 aIGlnArgLeu 38
| | | | | : : : : : |
375 TTGAAAGAGTC 385

```

**THIS PAGE BLANK (USP 11)**



```

seq_documentation_block:
LOCUS AL644594 645 bp mRNA linear EST 07-NOV-2001
DEFINITION AL644594 XGC-egg Silurana tropicalis cDNA clone L1E1d12 5', mRNA
sequence.
ACCESSION AL644594
VERSION AL644594.1 GI:16796719
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Silurana.
REFERENCE
1 (bases 1 to 645)
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
JOURNAL Unpublished (2001)
COMMENT
Contact: Huckle E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: L1E1d12.plc
Sequencing primer: PLC
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
FEATURES
source
1..645
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="L1E1d12"
/clone_lib="XGC-egg"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"
BASE COUNT 222 a 125 c 156 g 141 t 1 others
ORIGIN
alignment_scores:
Quality: 183.00 Length: 39
Ratio: 4.816 Gaps: 0
Percent Similarity: 97.436 Percent Identity: 92.308
alignment_block:
US-08-973-363-9 x AL644594 ..
Align seg 1/1 to: AL644594 from: 1 to: 645
1 TTELEUPROASPSAPSPROASPLYSLSPPROGINALYSGLINLEUCLNTH 17
|||||
456 ATTTTAAACGAGTATCCAGATAAGAGCCCAAGCCCAACAGCTACAGAC 505
17 TARGALAASTPYTLEULLLYSLEUDENUSNYSASPLEALAAARGLYSG 34
|||||
506 CAGAGCTGACTACCTCATTTAACTTCTCATTAAGAATCTGCTTAGCAAG 555
34 TUVAGLNAARGLEUTHR 39
|||||
556 AAGACAACAAGACTTCT 572
seq_name: gb_est1.BB155356
seq_documentation_block:
LOCUS BB155356 619 bp mRNA linear EST 18-OCT-2001
DEFINITION BB155356 RIKEN full-length enriched, 16 days neonate thymus Mus
musculus cDNA clone A130024L16 3' similar to L10410 Mouse
DNA-binding protein (CHD-1) mRNA, mRNA sequence.
ACCESSION BB155356
VERSION BB155356.2 GI:16268254
KEYWORDS EST.

```

SOURCE	ORGANISM
mus musculus	mouse mouse.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus; 1 (bases 1 to 619)	
REFERENCE	Atkawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Himoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,U., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakawa,D., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
JOURNAL	RIKEN Mouse ESTs (Atkawa,T., et al. 2001)
COMMENT	Unpublished (2001) On Jun 29, 2000 this sequence version replaced gi:1811286. Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) Wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanaki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,U., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shingawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Please visit our web site ( <a href="http://genome.gsc.riken.go.jp/">http://genome.gsc.riken.go.jp/</a> ) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Location/Qualifiers 1. 619 /organism="Mus musculus" /db_xref="taxon:10090" /clone="AI30024L16" /clone_id="RIKEN full-length enriched, 16 days neonate thymus" /tissue_type="thymus" /dev_stage="16 days neonate" /lab_host="DPH10B" /note="Site_1: Salt; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15' GAGGAGAGAGAGATCCAAAGGCTCTTTTCTTTTCTTTTNN 3'1, cDNA was transcribed by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by

cap-trapper cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAAGATTCGACGTAAATTAAATTTCCCGCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBlueScript KS(+) after bulk excision from Lambda

BASIS COUNT 194 a 118 c 161 g 146 t

ORIGIN

alignment\_scores:

	Quality:	182.00	Length:	41
Percent Similarity:	Ratio:	4.350	Gaps:	0
	Percent Identity:	97.361		85.366

Alignment\_block:  
US-08-973-363-9 x BB155356 ..

Align seq 1/1 to: BB155356 from: 1 to: 619

```
1 lileleproApsApProkAsPlyLysPProGlnAlaLysGlnLeucDnH 17
|||||
489 ATTCTTCCAGATCATCTGGATAAAAAACACACAGCAAAACAGTAGTAAC 538
17 TargAlaAPryLeuIlEtLyLeuLeuAnSykSPheAlaATyGly 34
|||||
539 CCGTCAGACACTACCCTCATCAACACTTGACAGAGACTCTGCAAAAGAG 588
34 luvaGlInArGLEuthrGlyAla 41
|||:::||
589 AGGCTCAGAGACTTGTGTGTGG 611
```

seq\_name: gb\_estl:BB461065

seg\_documentation\_block:

LOCUS BB461065 660 bp mRNA linear EST 25-OCT-2001

DEFINITION Mus musculus full-length cDNA, 12 days embryo spinal ganglion

Mus musculus cDNA clone D13070B13.3 similar to U10410 Mouse DNA-binding protein (Cbp-1) mRNA, mRNA sequence.

VERSION BB461065

KEYWORDS BB461065.2 GI:16426612

SOURCE EST.

ORGANISM house mouse.

REFERENCE 1

REFERENCE AUTHORS Eukaryotes Metazoa: Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 660)

Arikawa,T., Carrincci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Kouda,H., Hiramoto,K., Hoti,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasakhi,D., Shibata,K., Shinagawa,A., Shitaki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.

Riken Mouse ESTs (Arkawa,T., et al. 2001)

Unpublished (2001)

Title genome-resgasc\_riken.go.jp  
URL:http://genome.gsc.riken.go.jp/  
Carinacci,P., Shibata,Y., Hayashi,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

TITLE  
JOURNAL  
COMMENT

```

MATCHIKI,M., YONEDA,Y., ISHIKAWA,T., OZAWA,K., TANAKA,T., MATSUURA
,S., KAWAI,J., OKAZAKI,Y., MURAMATSU,M., INOUE,Y., KIRA,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1457-1471 (2000)
Konno,H., FUKUSHI,Y., SHIBATA,K., ITOH,M., CARINCI,P., SUGAHARA
,Y. and Hayashizaki,Y.
Computer based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., SHINAGAWA,A., SAITO,T., KIYOSAWA,H., YAMANAKA,I., AIZAWA
,K., FUKUDA,S., HATA,A., ITOH,M., KAWAI,J., SHIBATA,K. and
Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.ysc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Location/Qualifiers
1. 660
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D13007B13"
/clone_lib="RIKEN full-length enriched, 12 days embryo
spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site.1: SalI; Site.2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 15'
GAGGAGAGAGAGCGCGCACTCGAGATTTTGTGTTTNN 3'. cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence 15'
GAGGAGAGATCTTCGAGTTAAATTAATTAATTCCTCCGCCCC 3'. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
phluescript KS(+) after bulk excision from lambda FIC I."
BASE COUNT 211 a 130 c 168 g 151 t
ORIGIN
alignment_scores:
Quality: 182.00 Length: 41
Ratio: 4.550 Gaps: 0
Percent Similarity: 97.561 Percent Identity: 85.366
Alignment block:
US-08-973-363-9 x BB461065 ..
Align seg 1/1 to: BB461065 from: 1 to: 660
1 llaaepProaspasProaspLylysgProGlnAlaLysGlnLength 17
|||||
481 AATCTTCGAGATGATCTCTATTAATAAACACACAGAAAGAGTGTACCAAC 530
|||||
17 TATGAlaAspTyrLeuLeuLeuLeuAsnLysAspLeuAlaTargLysG 34
|||||
531 CGGTGACAGACTACCTCATCAACTACTTACGACAGATCTTGCAAAAAAG 580
|||||
34 lvaalGlnArgLeuThrGlyAla 41
|||||
581 AGGCTTACAGAGACTTGTGCTGG 603

```



FEATURES	Location/Qualifiers
source	1. 438 /organism="Mus musculus" /db_xref="taxon:10090" /clone="G930013K04" /clone_lib="RIKEN full-length enriched, mammary gland RCB-0527 Jy9-MC(B) cDNA" /tissue_type="mammary gland" /cell_line="RCB-0527 Jy9-MC(B)"
BASE COUNT	138 a 96 c 108 g 96 t
ORIGIN	
alignment_scores:	
Quality:	167.00 Length: 42
Ratio:	4.175 Gaps: 1
Percent Similarity:	95.238 Percent Identity: 80.952
alignment_block:	
US-08-973-363-9 x BB830730 ..	
Align seg 1/1 to: BB830730 from: 1 to: 438	
seq_name: gb_estl:AL601246	
seq_documentation_block:	
LOCUS	AL601246 430 bp mRNA linear EST 14-AUG-2001
DEFINITION	DKEZP3131040.F1 313 (synonym: hlcc2) Homo sapiens cDNA clone
ACCESSION	DKEZP3131040.5, mRNA sequence.
VERSION	AL601246
KEYWORDS	AL601246.1 GI:15164752
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 430) Bloeker, H., Boecher, M., Brandt, P., Mewes, W., Well, B. and Wiemann, S. EST (Bloeker, H., Boecher, M., Brandt, P., Mewes, H.W., Well, B. and Wiemann, S.) unpublished (1999)
JOURNAL	CONTACT: Bloeker, H
COMMENT	MIPS Am Klopferplatz 18a D-82152 Martinsried, Germany This is the 5' sequence of the clone insert clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de; Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project. No sl sequence available. This clone (DKEZP3131040) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcententrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES	Location/Qualifiers
source	1. 430 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="DKFZP3131040" /clone_lib="313 (synonym: hlcc2)"

FEATURES	/dev_stage="adult" /lab_host="PH10B" /note="Vector: pTriplex2; site_1: sfIIa; site_2: sfIIb; cDNA-collection"
BASE COUNT	163 a 81 c 86 g 100 t
ORIGIN	
alignment_scores:	
Quality:	166.00 Length: 41
Ratio:	4.256 Gaps: 1
Percent Similarity:	95.122 Percent Identity: 82.927
alignment_block:	
US-08-973-363-9 x AL601246 ..	
Align seg 1/1 to: AL601246 from: 1 to: 430	
seq_name: gb_estl:AL890775	
seq_documentation_block:	
LOCUS	AL890775 547 bp mRNA linear EST 07-MAR-2000
DEFINITION	wm95f11.x1 NCI-CGAP ut2 Homo sapiens cDNA clone IMAGE:2443725 3'
ACCESSION	similar to SW:CHD1_THUMAN_014646 CHROMODOMAIN1-HELICASE-DNA-BINDING
VERSION	AL890775
KEYWORDS	AL890775.1 GI:5595939
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 547) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index unpublished (1997) CONTACT: Robert Strausberg, Ph.D. Email: cgaaps-r@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bhbp/image/image.html Insert length: 1924 Std Error: 0.00 Seq primer: ~40up from Cllico High quality sequence stop: 418.
FEATURES	Location/Qualifiers
source	1. 547 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2443725" /clone_lib="NCI-CGAP Ut2" /tissue_type="moderately-differentiated endometrial adenocarcinoma, 3 pooled tumors" /lab_host="DH10B" /note="Organ: uterus; Vector: pCMV-SPORT6; site_1: SalI; site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

BASE COUNT	114 a	118 c	85 g	230 t
ORIGIN				

## alignment\_scores:

Quality:	166.00	Length:	41
Ratio:	4.256	Gaps:	1
Percent Similarity:	95.122	Percent Identity:	82.927

## alignment\_block:

US-08-973-363-9 x AI890775/rev ..

Align seg 1/1 to reverse of: AI890775 from: 1 to: 547

```
1 11eleuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnth 17
|||||
378 ATTCTTCAGATGATCCCGATTAACCAAGCAAAACAGTTGCAGAC 329
17 rArgAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysg 34
|||||
328 CCCTGCAGACTACCTCATCAATTAAGATCTTGCAGAAAAAG 279
34 luValGlnArgLeuThrGlyAla 41
|||||
278 AAGCT.....CTTCTGCTGCG 262
```

seq\_name: gb\_est1:AU125712

## seq\_documentation\_block:

LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.

ACCESSION AU125712 GI:10950428  
VERSION AU125712.1  
KEYWORDS EST.

## SOURCE

ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,T., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.

## TITLE

HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
Y., Sugano,S., Isogai,T.)

JOURNAL  
COMMENT Unpublished (2000)  
Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomicehri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

## FEATURES

source

1..866

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="NT2RM4002061"

/clone\_lib="NT2RM4"

/cell\_type="teratocarcinoma"

/cell\_line="NT2"

/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

## alignment\_scores:

Quality:	166.00	Length:	41
Ratio:	4.256	Gaps:	1
Percent Similarity:	95.122	Percent Identity:	82.927

## alignment\_block:

US-08-973-363-9 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

```
1 11eleuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnth 17
|||||
450 ATTCTTCAGATGATCCCGATTAACCAAGCAAAACAGTTGCAGAC 499
17 rArgAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeuAlaArgLysg 34
|||||
500 CCCTGCAGACTACCTCATCAATTAAGATCTTGCAGAAAAAG 549
34 luValGlnArgLeuThrGlyAla 41
|||||
550 AAGCT.....CTTCTGCTGCG 566
```

seq\_name: gb\_est2:BE895133

## seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.

ACCESSION BE895133 GI:10358221  
VERSION BE895133.1  
KEYWORDS EST.

## SOURCE

ORGANISM

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1028)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaphs-ri@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DTF  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LLM9753 row: h column: 16  
High quality sequence stop: 488.

FEATURES  
source

1..1028

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3921087"

/clone\_lib="NIH\_MGC\_72"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: pCMV-SPORT6; Site:1; Nci;  
Site:2; Sall; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."

BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

## alignment\_scores:

Quality:	166.00	Length:	41
Ratio:	4.256	Gaps:	1
Percent Similarity:	95.122	Percent Identity:	82.927

## alignment\_block:





Align seg 1/1 to: BF239967 from: 1 to: 821

```

6 ProaSplstysProglAlaLysGlnLeuGlnThrArgAlaAspTyrLe 22
  |||
3 CCGGATAAAAAACACAGCAAAACAGTTGCACACCGCTCACACACTRCT 52
  |||
22 uileuysleuLeuAsnLysAspleuAlaArgLysGluValGlnArgLeu 39
  |||
53 CATCAATATCTAGTAGAGACTCTGTGCAAAAAGAAAGCT.....CTTT 96
  |||
39 hrclyAla 41
  |||
97 CTGGTGGC 104

```

seq\_name: gb\_estl:AM996787

seq\_documentation\_block:

LOCUS AM996787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AM996787  
VERSION AM996787.1 GI:8257021

KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 337)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-QV3-BN0047-230  
200-102-d03&t3=2000-02-23&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 2

High quality sequence stop: 337.  
Location/Qualifiers

FEATURES

SOURCE

1..337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"

/note="Organ: breast.normal; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

ORIGIN

alignment\_scores: Quality: 129.50 Length: 35  
Ratio: 3.924 Gaps: 2  
Percent Similarity: 94.286 Percent Identity: 82.857

alignment\_block:

US-08-973-363-9 x AM996787/rev ...

Align seg 1/1 to reverse of: AM996787 from: 1 to: 337

```

1 lleuProaspaspProaSplstysProglAlaLysGlnLeuGlnThr 17
  |||
105 ATTCTTCAGATGATGCCGATTAATAAACACAGCAAAACAGTTGCAGAC 56
  |||
17 rargAlaAspTyrLeuileuys.leuLeuAsnLysAspleuAlaArgLys 33
  |||
55 CCGTGAGACTATCTCATCAATACTT...AGTAGAGATCTTGCAAAAAA 9
  |||
34 Glu 34
  |||
8 GAA 6

```

seq\_name: gb\_estl:AM997058

seq\_documentation\_block:

LOCUS AM997058 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AM997058  
VERSION AM997058.1 GI:8257292

KEYWORDS EST.  
SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 686)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=et2-QV3-BN0047-150  
400-152-c03&t3=2000-04-15&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 19

High quality sequence stop: 678.  
Location/Qualifiers

FEATURES

source

1..686  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"

/note="Organ: breast.normal; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 165 a 154 c 126 g 241 t

ORIGIN

## alignment\_scores:

Quality: 126.00 Length: 33  
Ratio: 3.818 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 87.879

## alignment\_block:

US-08-973-363-9 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

```
1 11leuProASPAPSPProASPlySLysProGlnAlaIysGlnleuGlnTh 17
|||||
127 ATTCTTCAGATGATGCCGATAA.AAACCAAGCAAAACATTGCAGAC 79
|||||
17 ATGAlaAspTYrIleuIleuIleuAsnIysAspIleuAlaArgLys 33
|||||
78 CCGTCAGACTACTCTCAATCTTAGAGATCTTGCAAAAGA 30
```

seq\_name: gp\_gss:CNS04JPY

## seq\_documentation\_block:

LOCUS CNS04JPY 1047 bp DNA linear GSS 21-MAY-2000  
DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone  
14P18 of library G from Tetradon nigroviridis, genomic survey  
sequence.  
ACCESSION AL293839.1 GI:8032419  
VERSION AL293839  
KEYWORDS GSS; genome survey sequence.  
SOURCE Tetradon nigroviridis.  
ORGANISM Tetradon nigroviridis  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetradon.

REFERENCE 1 (bases 1 to 1047)  
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetradon nigroviridis  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1047)  
Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using  
Tetradon nigroviridis DNA sequence  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 1047)  
Genoscope.

COMMENT Direct Submission  
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetradon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetradon>.

## FEATURES

source 1..1047  
/organism="Tetradon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="114P18"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C0BGL1ADH09LPI-end : T7"  
BASE COUNT 226 a 270 c 244 g 297 t 10 others  
ORIGIN

## alignment\_scores:

Quality: 124.50 Length: 40  
Ratio: 3.662 Gaps: 1  
Percent Similarity: 85.000 Percent Identity: 65.000

## alignment\_block:

US-08-973-363-9 x CNS04JPY/rev ..

Align seg 1/1 to reverse of: CNS04JPY from: 1 to: 1047

```
1 11leuProASPAPSPProASPlySLysProGlnAlaIysGlnleuGlnTh 17
|||||
465 ATTCTACAGATGATCCAGCAAGACCCAGTCACAGCAGCTACAGGC 416
|||||
17 ATGAlaAspTYrIleuIleuIleuAsnIysAspIleuAlaArgLys 34
|||||
415 CAGAGCTGAGATATCTCTCAAGCTGTGAAGAAAGAACAGACGACG 366
|||||
34 luvalGlnArgleuThrgly 40
|||||
365 AG...CAGTCCACACACAGA 349
```





KEYTNPBQIKQMRKNMIEVSKTFEDARKLHKYHAIKKROESQOHNDONISSNVN  
TRVIRNDPDRLEKETYNHDSRSDYSDDHLSQYHDHNRDRGDAYKKSDSKRPY  
SAFNGDHRDMDHKQDSRYSDSKRKLDDHRSRDRSRLNLEGNLKDSGSHDRSH  
SDRHSDDHSTSEYSHKSSRDYRHSMDQMHRASGSGPRSLDQSPYGSSPUG  
HRSPEHSSDHKSTPEHTWSRKT"  
4341..4604  
/gene="CHD-2"  
/note="short insert found in longer variant mRNA of CHD-2"

misc\_feature  
BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN

alignment\_scores:  
Quality: 445.00 Length: 88  
Ratio: 5.057 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AF004397 ..

Align seg 1/1 to: AF004397 from: 1 to: 6872

1 ASGGUULIEVALSERVALLYSHISLEUHSLSYSLYSLIELYSTHGLU 17  
|||||  
4341 GATGAGATTGTTTCAGTGAAACATCTACATAAAAAATAAAAAACAGAAAA 4390  
17 SGLUASNGUGLULYSGLUASPLYSGLULEULYSGLULYSASPASNL 34  
|||||  
4391 AGAANAATGAGAAAAAGCCGACCCAGATTTGTGTATAAGAAAGAAAGCTG 4440  
34 LUGLULYSARGGLUTHTLYSGLUASNLYSGLULYSARGGLULEULYSAR 50  
|||||  
4441 AAGAAAAAGAGAGACAAAGAGAAAGAAAAATAAAAAGGAAATGAAAAAG 4490  
51 GLULYSGLULYSGLULYSGLULYSGLULYSGLULYSASPASNL 67  
|||||  
4491 GAGAAAAAGAAAAAGAGATAGAAAGAAATAAAAAGAAAGATTAATA 4540  
67 SGLULYSARGGLUASNLYSVALLYSGLUSETHRCGLNLSGLULYSGLULY 84  
|||||  
4541 AGAAAAAGAGAAAAACAAAGTAAAAAGAAATCCACACAGAAAAAGAAAG 4590  
84 ALYLSGLULULYS 88  
|||||  
4591 TCAAGAGAGAGAGAG 4604

seq\_name: gb\_pat:A58693

seq\_documentation\_block:

LOCUS A58693 265 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 12 from Patent WO9639505.  
ACCESSION A58693  
VERSION A58693.1 GI:3714251

KEYWORDS  
SOURCE  
ORGANISM

unidentified.  
unclassified.

REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 265)  
Griffiths,R. and Tiwari,B.  
AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS

JOURNAL  
PATENT: WO 9639505-A 12 12-DEC-1996;  
ISIS INNOVATION (GB)  
Other publication AU 5906996 961224.

COMMENT  
FEATURES  
source  
1..265  
Location/Qualifiers

location="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 158 a 16 c 61 g 30 t  
ORIGIN

alignment\_scores:

Quality: 432.00 Length: 89  
Ratio: 4.909 Gaps: 1  
Percent Similarity: 98.876 Percent Identity: 98.876

alignment\_block:

US-08-973-363-11 x A58693 ..

Align seg 1/1 to: A58693 from: 1 to: 265

1 ASGGUULIEVALSERVALLYSHISLEUHSLSYSLYSLIELYSTHGLU 17  
|||||  
1 GATGAGATTGTTTCAGTGAAACATCTACATAAAAAATAAAAAACAGAAAA 50  
17 SGLUASNGUGLULYSGLUASPLYSGLULEULYSGLULYSASPASNL 33  
|||||  
51 AAGAAAAATGAGAAAAAGCCGACCCAGATTTGTGTATAAGAAAGAAAGCT 100  
34 LUGLULYSARGGLUTHTLYSGLUASNLYSGLULYSARGGLULEULYSAR 50  
|||||  
101 GAGAAAAAGAGAGACACAAAGAGAAAGAAAAATAAAAAAGGAAATGAAAA 150  
50 GGLULYSGLULYSGLULYSGLULYSGLULYSGLULYSASPASNL 67  
|||||  
151 GGAGAAAAAGAAAAAGAGATAGAAAGAAATTAAGAAAAAGAAAGATTA 200  
67 YSGLULYSARGGLUASNLYSVALLYSGLUSETHRCGLNLSGLULYSGLULY 83  
|||||  
201 AAGAAAAAGAGAAAAACAAAGTAAAAAGAAATCCACACAGAAAAAGAA 250  
84 VALYLSGLULULYS 88  
|||||  
251 GTCAAGAGAGAGAGAG 265

seq\_name: gb\_dr:AC092372

seq\_documentation\_block:

LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
ACCESSION AC092372  
VERSION AC092372.3 GI:17402768

KEYWORDS  
SOURCE

human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL  
TITLE  
DOE Joint Genome Institute and Stanford Human Genome Center.

REFERENCE  
AUTHORS  
TITLE  
Direct Submission  
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

JOURNAL  
TITLE  
Direct Submission  
Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA

COMMENT  
FEATURES  
source  
1..101220  
Location/Qualifiers

location="Homo sapiens"  
/db\_xref="taxon:9606"

location="unidentified"  
/db\_xref="taxon:32644"

location="unidentified"  
/db\_xref="taxon:32644"

location="unidentified"  
/db\_xref="taxon:32644"

location="unidentified"  
/db\_xref="taxon:32644"

location="unidentified"  
/db\_xref="taxon:32644"

location="unidentified"  
/db\_xref="taxon:32644"

location="unidentified"  
/db\_xref="taxon:32644"

location="unidentified"  
/db\_xref="taxon:32644"

location="unidentified"  
/db\_xref="taxon:32644"

seq\_name: gb\_pr:AC012624

seq\_documentation\_block:

LOCUS AC012624 13365 bp DNA linear PRI 21-JUL-2001

DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.

ACCESSION AC012624

VERSION AC012624.6 GI:14993679

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 13365)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Unpublished

JOURNAL Direct Submission

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (31-Oct-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 2 (bases 1 to 13365)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 13365)

AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.

Location/Qualifiers

BASE COUNT 34122 a 18862 c 17827 g 30409 t

ORIGIN

Alignment-scores:

Quality: 356.00 Length: 88

Ratio: 4.395 Gaps: 0

Percent Similarity: 92.045 Percent Identity: 78.409

alignment\_block:

US-08-973-363-11 x AC029372/rev ..

Align seg 1/1 to reverse of: AC029372 from: 1 to: 101220

1 ASGTLTLEVALSERVALLYSRHSLSLEUHSLSYSLYSLLEYSYTHGGLU 17

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

25410 GATGAGATCAGTCTGTGTAACATCCAAATATMAAAATTAACAGAAAG 25361

17 SGLUASNGUGLULYSRPGIUPROASPILEGLYILEYLSYSGUAIAG 34

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

25360 AGACACTGAAGAAAACCTGACGACATCTTATATTAAGAAAGAACAG 25311

34 IUGLULYSARGGLUThrySGULYSGLUASNLYSARGGLULEULYSARG 50

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

25310 AAGAAAAGGAGACCAAAAGAAAGAAAGAAAGAAAGAAAGCTTAAAG 25261

51 GLULYSGLULYSGLULYSGLULYSGLULYSGLULYSGLULYSASP 67

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

25260 GAGATTAAGAAAAGAGATTAAGAAAGATTAAGAAAGAAAGAAAGCTTAA 25211

67 SGLULYSARGGLUASNLYSVALYSGLUSERThryGLULYSGLULY 84

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

25210 AGAAAAGAGAAAAGAAAGATTAAGAAAGCTATACAGAAAAGAAAGCA 25161

84 alysglululs 88

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

25160 TAAAGGAAAGAAAG 25147

source 1. 134365

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="CTD-2082117"

BASE COUNT 40414 a 24497 c 25503 g 43951 t

ORIGIN

Alignment-scores:

Quality: 356.00 Length: 88

Ratio: 4.395 Gaps: 0

Percent Similarity: 92.045 Percent Identity: 78.409

alignment\_block:

US-08-973-363-11 x AC012624 ..

Align seg 1/1 to: AC012624 from: 1 to: 134365

1 ASGTLTLEVALSERVALLYSRHSLSLEUHSLSYSLYSLLEYSYTHGGLU 17

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

119773 GATGAGATCAGTCTGTGTAACATCCAAATATMAAAATTAACAGAAAG 119822

17 SGLUASNGUGLULYSRPGIUPROASPILEGLYILEYLSYSGUAIAG 34

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

119823 AGACACTGAAGAAAACCTGACGACATCTTATATTAAGAAAGAACAG 119872

34 IUGLULYSARGGLUThrySGULYSGLUASNLYSARGGLULEULYSARG 50

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

119873 AAGAAAAGGAGACCAAAAGAAAGAAAGAAAGAAAGAAAGCTTAAAG 119922

51 GLULYSGLULYSGLULYSGLULYSGLULYSGLULYSGLULYSASP 67

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

119923 GAGATTAAGAAAAGAGATTAAGAAAGATTAAGAAAGAAAGAAAGCTTAA 119972

67 SGLULYSARGGLUASNLYSVALYSGLUSERThryGLULYSGLULY 84

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

119973 AGAAAAGAGAAAAGAAAGATTAAGAAAGCTATACAGAAAAGAAAGCA 120022

84 alysglululs 88

||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

120023 TAAAGGAAAGAAAG 120036

seq\_name: gb\_htg:AC021449

seq\_documentation\_block:

LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000

DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered pieces.

ACCESSION AC021449

VERSION AC021449.3 GI:10047806

KEYWORDS HTG; HTGS; PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 143079)

AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE Unpublished

JOURNAL 2 (bases 1 to 143079)

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F., Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A., Chongslavsky,L., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domingo,M., Doyle,M., Faganer,D., Ferreira,P., Fitzhugh,N., Forrest,C., Gage,D., Galagan,D., Gardy,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczeky,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N., McEwan,P., McCut,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,

**TITLE** Direct Submission  
**JOURNAL** Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Sep 10, 2000 this sequence version replaced gi:17407963.

Plattie,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.  
Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Rhomon,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wymann,D., Ye,W.,J.  
Zimmer,A. and Zody,M.

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIPR  
 Web site: <http://www.seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 Project Information

```
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 134745 bases at least Q40
Consensus quality: 139227 bases at least Q30
Consensus quality: 140814 bases at least Q20
Insert size: 144000; agarose-fp
Insert size: 142179; sum-of-ctrls
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-ctrls
```

NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

*	1	38820:	confly of 38820 bp in length
*	38621	38920:	gap of 100 bp
*	38921	40411:	confly of 1491 bp in length
*	40412	40511:	gap of 100 bp
*	40512	43279:	confly of 2768 bp in length
*	43280	43379:	gap of 100 bp
*	43380	46905:	confly of 3526 bp in length
*	46906	47005:	gap of 100 bp
*	47006	51830:	confly of 4825 bp in length
*	51831	51930:	gap of 100 bp
*	51931	62619:	confly of 10683 bp in length
*	62620	62719:	gap of 100 bp
*	62720	75408:	confly of 12683 bp in length
*	75409	75508:	gap of 100 bp
*	75509	92516:	confly of 17008 bp in length
*	92517	92616:	gap of 100 bp
*	92617	106409:	confly of 13793 bp in length
*	106410	106509:	gap of 100 bp
*	106510	143079:	confly of 36570 bp in length

```

FEATURES
  source
    Location/Qualifiers
      1..143079
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RP11-58M12"
        /clone_lib="RPCT-11 Human Male BAC
          1..38820
            misc_feature
              /note="assembly-fragment
                clone_end:SP6
                vector_side:left"
                38921..40411
                  /note="assembly-fragment"
                  40512..43779
                    misc_feature
                      /note="assembly-fragment"

```

misc_feature	43380_..46905	/note="assembly-fragment"
misc_feature	47006_..51830	/note="assembly-fragment"
misc_feature	51931_..62619	/note="assembly-fragment"
misc_feature	62720_..75408	/note="assembly-fragment"
misc_feature	75509_..92516	/note="assembly-fragment"
misc_feature	92617_..106409	/note="assembly-fragment"
misc_feature	106510_..143079	/note="assembly-fragment"

```
alignment_scores:
  Quality: 356.00      length: 88
  Ratio: 4.395        Gaps: 0
Percent Similarity: 92.045      Percent Identity: 78.409
```

alignment\_block:  
US-08-973-363-11 x AC021449

Align seg 1/1 to: AC021449 from: 1 to: 143079

1 AspgluilevalserValylshisleuHisllyslyslieleysthrgIuly 17  
||||||| ||||||||| :::||||| ||||||||| :::  
117669 GATGAGATCAGTCTGTGAACATCCAAATAAAAAATTAACAGAAG 117718

```

17  sgluasngluglulysprogluPraspilleglyilelysylsgluAlag 34
   :::::|||||:::|||||
117719 AGACAGTGAAGAAAACTGAGCCAGACTGTTATATATAAGGAAGAACCCAG 117768

```

34 IuGIuLySaRgCIuThrLySGluYsGIuAsnLySaRgCIuLeuLySaRg 50  
|||||:::|||||:::|||||  
117769 AAGAAAGAGAGCGAACAAGAAAGAGAGAGATAAAAAAGACTTAAAGC 117818

**51** GLGLSLSLSGIATSGLVASPLSLSGLELSGIADLSASPANLy **67**

||| |||||||||||||||||||||:::||||| |

**117819** GAGATAAAGCAAAAAGAGGATACGAACATATTAAAGCAAAGAATTTTAA **117868**

67 sGIuLyArGtIuaSnLySvAlLyScLuSeRThrGlInLySGlUy 84  
|||||  
117869 AGAAAGAGACAAACCAAGTAAAGAGCTATACAGAGAAGAAAAAGACA 117918

```

      84  allysGluGluys  88
          ::| | | | | | | |
117919  TAAAGGAGAAAG 117932

```

seq\_name: gb\_htg:AC008531

```
seq_documentation_block:
  docno: 10000531
```

seq:AC008531	145659 bp	DNA	linear	HTG 14-FEB-2001
LOCUS AC008531				
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE.				
7 ordered pieces.				

ACCESSION	AC008531
VERSION	AC008531.3
KEYWORDS	GI:12630078
SOURCE	HTG; HTGS_PHASE2; HTGS_DRAFT human.

1  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 145659)  
REFERENCE  
AUTHORS  
DOE Joint Genome Institute.

TITLE	Sequencing of Human Chromosome 5
REFERENCE	2 (bases 1 to 145659)
JOURNAL	Unpublished
AUTHORS	DOE Joint Genome Institute.





*	9097	gap of unknown length
*	9197	10260: contig of 1064 bp in length
*	10261	10360: gap of unknown length
*	10361	12450: contig of 2100 bp in length
*	12461	12560: gap of unknown length
*	12561	14511: contig of 2051 bp in length
*	14612	14711: gap of unknown length
*	14712	16381: contig of 1670 bp in length
*	16382	17991: gap of unknown length
*	16482	17968: contig of 1487 bp in length
*	17969	18068: gap of unknown length
*	18069	20454: contig of 2366 bp in length
*	20455	20554: gap of unknown length
*	20535	22351: contig of 2981 bp in length
*	23316	23615: gap of unknown length
*	23616	27553: contig of 3948 bp in length
*	27564	27663: gap of unknown length
*	27664	30987: contig of 3324 bp in length
*	30988	31087: gap of unknown length
*	31088	36055: contig of 4978 bp in length
*	36056	36165: gap of unknown length
*	36166	40978: contig of 4813 bp in length
*	40979	41078: gap of unknown length
*	41079	45663: contig of 4585 bp in length
*	45664	45763: gap of unknown length
*	45764	51745: contig of 5982 bp in length
*	51746	51845: gap of unknown length
*	51846	57359: contig of 5514 bp in length
*	57360	57369: gap of unknown length
*	57460	67881: contig of 10422 bp in length
*	67882	67991: gap of unknown length
*	67982	74132: contig of 6151 bp in length
*	74133	74232: gap of unknown length
*	74233	79795: contig of 5563 bp in length
*	79796	79885: gap of unknown length
*	79896	87511: contig of 7616 bp in length
*	87512	87611: gap of unknown length
*	87612	92791: contig of 5180 bp in length
*	92792	92891: gap of unknown length
*	92892	102794: contig of 9903 bp in length
*	102795	102894: gap of unknown length
*	102895	110866: contig of 7972 bp in length
*	110867	110966: gap of unknown length
*	110967	117571: contig of 6605 bp in length
*	117572	117671: gap of unknown length
*	117672	123788: contig of 6067 bp in length
*	123789	123883: gap of unknown length
*	123839	130683: contig of 6745 bp in length
*	130684	130683: gap of unknown length
*	130684	141544: contig of 10861 bp in length
*	141545	141644: gap of unknown length
*	141645	169109: contig of 27465 bp in length
*	169110	gap of unknown length
*	169210	193446: contig of 24237 bp in length

BASE COUNT	57571	a	35252	c	37061	g	60360	t	3202	others
ORIGIN										

```
alignment_scores:
  Quality: 356.00
  Ratio: 4.395
  Percent Similarity: 92.045
  Length: 88
  Gaps: 0
  Percent Identity: 78.409
```

```
alignment_block:
US-08-973-363-11 x AC091946
```

Align seg 1/1 to: AC091946 from: 1 to: 193446

1 AsprgluIeValSerValLysHisLeuHisLysLysIleLysThrGlu 17

89180 GATGAGATCAGTCTCTGTGAACATCCAAATAAAAAATTAAAAACAGAAG 89229

17 SGluAsnGluGlyProGluProAspIleGlyIleLysLysGluAlaG 34

89230 AGACAGTGAAGAAAAACCTGAGCCAGATGTTTATATATAAGAGAAGAACAG 89279

34 IuGIuLysArgGIuThrLysGIuLysGIuAsnLysArgGIuLeuLysArg 50

89280 AAGAAAAGAGCGAAGCAAAAGAAAGGAGATAAAAAAGAACTTAAAGG 89329

```
31 GtUylstysGULySGIwasplYslYSglubeulYsgILlYsaspassnly b/
```

[illegible]

89380 AGAAAAAGAGAAAACAAGTAAAGAAGCTATACAGAAAGAAAGACA 89429

84 allysclugluls 88

89430 TAAAGGAAGAAAG 89443

seq\_name: gb\_pr:AC026778

seq\_documentation\_block:

LOCUS	AC026778	195433 bp	DNA	linear	PRI 01-JUN-2001
DEFINITION	Homo sapiens chromosome 5 clone CTC-42811, complete sequence.				

AC026778	GI:14277282
AC026778.4	

SOURCE human

## DISCUSSION

REFERENCE 1 (bases 1 to 195433)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**AUTHORS** DOE Joint Genome Institute and Stanford Human Genome Center.

Unpublished  
JOURNAL

2 (bases 1 to 193433)  
DOE Joint Genome Institute.

TITLE	Direct Submission	Submitted (24-MAR-2000)	Production Sequencing Facility	DOE Joint
JOURNAL				

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

**AUTHORS** DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission  
 JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell

COMMENT On Jun 1 2001 this sequence replaced ri.13677045  
Drive, Walnut Creek, CA 94598, USA

Draft Sequence Produced by DOE Joint Genome Institute

Finishing Completed at Stanford Human Genome Center

Quality: Phrap Quality >=40 99.9% of Sequence:  
www.singc.stanford.edu

Estimated Total Number of Errors is 0.2.  
STS Content:

WI-13675 G23101

SHGC-103595 G57841.

```
FEATURES
source      1. 195433
location/qualifiers
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

/chromosome=5

BASE COUNT	62762 a	37302 c	37040 g	58329 t

DATE CONT  
ORIGIN



```
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 435334
Center clone name: RPCI-11_75H1
-----
Summary Statistics
Consensus quality: 229677 bases at least Q40
Consensus quality: 256163 bases at least Q30
Consensus quality: 260799 bases at least Q20
Estimated insert size: 174820; agarose-1p estimation
Estimated insert size: 271581; sum-of-contigs estimation
Quality coverage: 8.62 in Q20 bases; agarose-1p estimation
Quality coverage: 5.55 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 47 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1061: contig of 1061 bp in length
1062
1161: gap of unknown length
1162
2827: contig of 1666 bp in length
2828
2927: gap of unknown length
2928
4227: contig of 1300 bp in length
4228
4327: gap of unknown length
4328
5618: contig of 1291 bp in length
5619
5718: gap of unknown length
5719
6983: contig of 1265 bp in length
6984
7083: gap of unknown length
7084
8422: contig of 1339 bp in length
8423
8523: gap of unknown length
8523
9771: contig of 1249 bp in length
9772
9871: gap of unknown length
9872
11198: contig of 1327 bp in length
11199
12128: gap of unknown length
12129
12414: contig of 1116 bp in length
12415
12514: gap of unknown length
12515
13806: contig of 1292 bp in length
13807
13906: gap of unknown length
13907
15360: contig of 1454 bp in length
15361
15460: gap of unknown length
15461
16667: contig of 1207 bp in length
16668
16767: gap of unknown length
16768
18390: contig of 1623 bp in length
18391
18490: gap of unknown length
18491
19541: contig of 1051 bp in length
19542
19641: gap of unknown length
19642
21156: contig of 1515 bp in length
21157
21256: gap of unknown length
21257
22448: contig of 1192 bp in length
22449
22548: gap of unknown length
22549
24499: contig of 1951 bp in length
24500
24599: gap of unknown length
24600
26384: contig of 1785 bp in length
26385
26484: gap of unknown length
26485
27883: contig of 1399 bp in length
27884
27983: gap of unknown length
27984
30386: contig of 2403 bp in length
30387
30486: gap of unknown length
30487
32480: contig of 1994 bp in length
32481
32580: gap of unknown length
32581
35441: contig of 2861 bp in length
35442
35541: gap of unknown length
35542
37800: contig of 2259 bp in length
37801
37900: gap of unknown length
37901
40587: contig of 2687 bp in length
40588
40687: gap of unknown length
40688
44159: contig of 3472 bp in length
44160
44259: gap of unknown length
44260
46336: contig of 2377 bp in length
```

```

* 46637 46736: gap of unknown length
* 46737 50082: contig of 3346 bp in length
* 50083 50182: gap of unknown length
* 50183 53988: contig of 3806 bp in length
* 53989 54088: gap of unknown length
* 54089 56592: contig of 2504 bp in length
* 56593 56692: gap of unknown length
* 56693 61353: contig of 4660 bp in length
* 61353 61453: gap of unknown length
* 61453 64254: contig of 2802 bp in length
* 64255 64354: gap of unknown length
* 64355 68825: contig of 4471 bp in length
* 68826 68925: gap of unknown length
* 68926 74393: contig of 5468 bp in length
* 74394 74493: gap of unknown length
* 74494 80561: contig of 6068 bp in length
* 80562 80661: gap of unknown length
* 80662 87626: contig of 6965 bp in length
* 87627 87726: gap of unknown length
* 87727 93600: contig of 5874 bp in length
* 93601 93700: gap of unknown length
* 93701 102024: contig of 8324 bp in length
* 102025 102124: gap of unknown length
* 102125 109905: contig of 7781 bp in length
* 109906 110005: gap of unknown length
* 110006 116737: contig of 6732 bp in length
* 116738 116837: gap of unknown length
* 116838 126797: contig of 9960 bp in length
* 126798 126897: gap of unknown length
* 126898 135114: contig of 8217 bp in length
* 135115 135214: gap of unknown length
* 135215 146713: contig of 11499 bp in length
* 146714 146813: gap of unknown length
* 146814 157894: contig of 11081 bp in length
* 157895 157994: gap of unknown length
* 157995 171752: contig of 13758 bp in length
* 171753 171852: gap of unknown length
* 171853 201687: contig of 29835 bp in length
* 201688 231268: gap of unknown length
* 231269 231368: contig of 29481 bp in length
* 231369 276181: contig of 44813 bp in length.
* 276181
* 1..276181
* location/Qualifiers
*
FEATURES
source
1..276181
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-75H1"
/clone_lib="RPCI human BAC library 11"
BASE COUNT 86859 a 51769 c 50026 g 82854 t 4673 others
ORIGIN
Alignment scores:
Quality: 356.00 Length: 88
Ratio: 4.395 Gaps: 0
Percent Similarity: 92.045 Percent Identity: 78.409
alignment_block:
US-08-973-363-11 x AC092382/rev ..
Align seg 1/1 to reverse of: AC092382 from: 1 to: 276181
1 AaGluUlleValSerValIysHisIeuHsIysIysIleIysThGluIy 17
||||||| | | | | | | | | | | | | | | | | | | | | | | |
212172 GATGAGATCAGTCTGTGTGAACATCAATATAAAAAATTAAAG 212123
17 sGluAsngIuGluIysProGluProAspIleGlyIleIysIysGluAaIag 34
::: | | | | | | | | | | | | | | | | | | | | | | | |
212122 AGACAGTGAAGAAAAACCGACCGAGTGTATATTAAGAAAGAACG 212073
34 LuGluIysArgGluThrIysGluIysGluAsnIysArgGluLeuIysArg 50
| | | | | | | | | | | | | | | | | | | | | | | | | | |
```







/protein\_id="A058053.1"  
/db\_xref="GI:2337969"  
/translation="MSQNSNSNDSPKRRKRYKMKDLTEBOKERRRSINRASKNFK  
RRRIPEOQKGLINLKYEENSLRCOYERKRDRLRLRWLNHKTCTONNYNGPE  
PRYKVENSLMQCATAFNLNDQOTTNINLIPEYVSGNNNTNGFAATATLHTNCKE  
TLANNINFEAKLNCLEVLPSFTSALDLSIDMWNINL"  
3492..5453  
/note="A3: AHV-sema, similar to Vaccinia A39"

CDS

/codon\_start=1  
/evidence=not\_experimental  
/product="semaphorin homolog"  
/protein\_id="A058054.1"  
/db\_xref="GI:2337970"  
/translation="MAYLNATVSKPVLSLSLKVKLFKEHCGEGCGLGLETVIH  
PAMGTLCAVIRILMILSATTAASKRFDKRLIVNLTDGGGHRFGPQPHVLEH  
SLNSDVGNGNNTIYLFDFHSSNALINITSPTNHRSLSTCENFTTLIHND  
GLACGNSKQSPSCMLNNITTOPLKCLAPSPSSGNLVFDONDYITNLKYS  
LGSJHFRRIAGVVELTSDTAHRRPOVATVHKNESYDKITFFEOBNSHDEKO  
FPHTPRVGVQSSDQGESLSVYKMTTELKARLACVDYDGRIRVNLQDIFIMQAP  
ENSMETELIYGLFSLPMNFSACVFTVKDIDHFKTSLKNYHKLPTPPGQCKMH  
OHVTEFQVADRYREYADPVYOKNNAMFIIOSKYITKLAVRYEGGVNATIFY  
LTIKGIHIVYREDSNSTALILEINFORPAPIONITLNTNKLTVNSEMEVS  
EVLIDCSYGNCCFCFMSRDPCTIYNTCSFKQVSEYETGPARNTISEMGDIY  
APTVKHQVSIPLNSYLSCPAVSNMADFWTKDFTFKCHVKTKNCKILLIANS  
TITNGIHVCNMKEDSVTKLEVNITLM"  
5590..5597  
5732..6097  
/note="ORF containing putative signal peptide; detected by  
5'RACE-PCR"

polyA\_signal  
CDS

/codon\_start=1  
/evidence=experimental  
/product="A4"  
/protein\_id="A058055.1"  
/db\_xref="GI:2337971"  
/translation="MCAQLYHHLTILIASYIIFVNAAPTLXEDDEEDNSTEYDF  
LSDSROFPRNSSCNTREARTRTLERALKRLQIAGITNETDELPYPTMTT  
PLETPTDTSPPVPSAIP"  
6123..6128  
/evidence=not\_experimental  
6333..10442  
/note="ORF03: similar to H. salmieri and EHV2 ORF3, similar  
to ORF75"

polyA\_signal  
CDS

/codon\_start=1  
/evidence=not\_experimental  
/product="tegument protein"  
/protein\_id="A058056.1"  
/db\_xref="GI:2337972"  
/translation="MITRELQLLATYTAEPALSTAAIAIRSTPGDQNYITQDAVL  
VFFTSRPRRLREHQLKIEITLFTKALRTFDEQOYLFIPLSALSSSTFYVGPDIINL  
PTSQSNELTALIQTRADAREKVENVERIOHCRLLFFSGSGEQLNHTHYALREILCGD  
LTLYOFRNELFNSSLDYFATIYPRFLKDKRHHVYGDINLTGAASYNNAASPPT  
AVYDSIFRIPTHVYRVIDRWEIRVAVPVGFGHSGVRFROHGAAYNCSLTNSK  
GATGCSLOKMITSFISOPTGOLGILTEPMCGMOELSATIOMLTRKANVYEAFAK  
TLNAGVPLVQGVFVTSPTLRKRAAFGTSLSITGPIITLPPATETIDAMERRR  
OULVEVGPDIYCGKGLHNPVNIANSNGRHAHILHOLQOLMITPEGVHITICDM  
QNEKVTMRKIAACGDLGISMARSRLPAHTLRLKKNFRNLNINRITIKNSLWGL  
SAALAVVADNEODVSKMDIALSGGCFHILGNTLTPSNTIIVSDKQOYEIADI  
OYKMHMPKQSEGETEDPLADMSNIOLKFNLDYEDLLQVLRHPVGGKANHVVH  
DRCGNGHIAQGVGPDIPLCDPSTVHNLVGDGIREGMEVPRVMAADRVARLLI  
ERYSTPGDIDPATIANNIGLTHYHPSEQVYVESKRYCNGIGICKTTFTQDRLP  
GTLIAVESCTNCTILGVENTEELIGLSISVEGHIYREHVSIMAAADFCSSNMF  
GCVNSAENGNCLRSVAVANAPCVPGPSLPYFKPGSALIRVNLHNEHFLSGGI  
CMAASIGASEFTPTPSOLRNLQOELIYKARNIALSGHVSIDGLICAVCEMFG  
GLSARLIIDHEDEPVPLESETPGLVEVNLIDVAAIIRANLYNEGYOIEHLYVS  
DRETFVHONTOLLISVPSRLKHNMTLFESKVDLLIYKEDVLPBETSSEGYEHLIYD  
PYTIISSOTTRPVNLVHLPGCGYPDALALALNSGSPDTVYIPGCKIYHNREDA  
GSPPTADDFIAGIIVLIGSSNIDSDVGSTIRQWLNVRQVINDVRAKAFSGTALAI  
GOLACRLFAITKAGIDAGSOQTFLLPNASRYESRWLNFKIPEDEKAAAFDRLAC  
VLPCWOGITLGFSSHNITPFGDLETRQOAAATNGPLVOSGAPAREPLAPTEAEHY  
AGCSEDEGRLALLPDCLAFTNMOQHNOTGQGGELPVSPPKIMYRLIYNMSKFPQ  
HYRSLRTNLRHFNFTQQLDFPHADRRAVPHDNQYITPMAAPSO"  
11134..11139  
/evidence=not\_experimental

polyA\_signal

CDS

11260..14643  
/note="ORF06"  
/codon\_start=1  
/evidence=not\_experimental  
/product="major ss DNA binding protein"  
/protein\_id="A058057.1"  
/db\_xref="GI:2337973"  
/translation="MALKHQLNHVEEDNLSKAPIGCGFIITYPTEHNFKEISLIG  
DKYAEAGFSLPLIHGVTEAFVPPNKAAYKKIDMTYTSVKISTSTYNAIIPHNK  
FESTFSGRGLSKKACDIFGVYPTPILEGESTDSDICPPWQEKDALMAYVITBG  
FKERHLIGKLLILKSDMHSVINKTEYRILPDLDELFKSSLRLRLYPAVSEVILY  
TLVYLSAQSLRVHNASLVEAIQEOPHDKYAKLVSEYREYLAIVAGADTLLMVID  
AAVAELGLSYSLSEFFEPQKRTQVODYSNDAFSCETSDREALSALMLOAIIH  
AOLPSTSIYYNRRVAAOAPIPNKSEYVNNYNYVOHGLANLCEETLEFDGSPATFG  
APASLIDGSEFTIOLHAAAFSPNLLABWCYIOLGOCALDELTNANVTTEVGSAA  
NSPVCSLCSQCCVCYCNITFYALKOREPVLQSGRRDPYVITGTTNVEDLFGNF  
ASFRKDQEDONQTEETRYTQWLNQTLTEKLAAGLVSDPAVDEBAGSGSGSLNKEF  
VRFPSDLSLVDAEAKETMTMKNVNFESIKGVSHYOVKNNTYWAQCSLMNL  
YYSILITIIIDIALPISTYESNPAGQYKPNMILKLYOTLNTFRSEFIDGVITG  
TEKMYVABQSPDFPDATNNMSPYKQVRLARAQVLAALKNIKVKNILRSGTSM  
SEHYQNAFLKTANRQNYILAGPYVFLANSFHQDLRPNLKSCLTYLMSNCKKQIOPC  
VPGVSALAKKRFYSYIINNSKQFEEVNMADVPDYVYTKAKOLNALLPACQTOFY  
AVTHTSIFPVQETCALETPHVLGTSVDSVEDYVNNVQMLKALIVNSSLRESANIL  
RSRPVTLIPVYVKNRYTGIAQNAQLFQSANLGYPMRGVKNLGLDELFKKQONSYMR  
KKYLFMTPLVGNLKPSTYHQGTAPELTKRTIOSLIEQDQADEVLAIRVCELVSL  
GACCADLTLDIOFYLGSYGMFSENIETEKIDQIRELVGPWTHMAESVLSKSGTCETDE  
VOEVAPEEBOVKRLSMDSHSGKVVGGKKRKIATMFDDLDL"  
14725..16767  
/note="ORF07"

CDS

/codon\_start=1  
/evidence=not\_experimental  
/product="transport protein"  
/protein\_id="A058058.1"  
/db\_xref="GI:2337974"  
/translation="MGQYLAALYQIYGLCDLVSYEPCKPTSLCTKTADACNKH  
IHEVSAKSLQONSLEKALSTELSHLEMLKRLFFVTHALDNPTYSKLSVGL  
CDLHAKRLNWQFVNEGILEVNLILINDIERLSLQVFCYALNDNPTYSKLSVGL  
QLRGILNWPQIDYITSSCLECVLEFVSYPNGOETLNEILNHNHNLHYERPPPI  
KGLFSELOMLGLKVAHATDITIOSQGHVAVGOESLVAUYLKATITNNPKOYALEN  
LIYWSGONQSPDSGVKCSSELKIMSENDELQYRKLNNGEPGPPHDSPOGTEL  
LRGGIJSFTHDITALKODCSNTPKOPFLPVAARONLFRLSNIIUYGEVPPKP  
KOTBSALKTCDQSDASKNOYLOAELEKRAYLKLKSEGRKLOACLSHTEBNLSQ  
SLKIWGSVYVQKQATLLNHLFLRQSWTQAPLSPVNSGSEQFNSKLSLYVKS  
SREYLSLTLRLHLPFALLTGPLTTOBGLFPPSPNOLAHCELAHFMHOKMLIEMKP  
HSETPDMICSNFENFEYTIHETDNGVQECWKILRELIVSVALYNTWEKNLICITRD  
TMCPTACSSGIRKGLVYTESHAPLILVYSNKKWIFRKDIYALLAHMQLANNAHR"  
15048..15053  
/evidence=not\_experimental

polyA\_signal  
CDS

/codon\_start=1  
/note="ORF08: similar to EBV BALF4, CMV UL55, HSV UL27"  
/codon\_start=1  
/evidence=not\_experimental  
/product="glycoprotein B"  
/protein\_id="A058059.1"  
/db\_xref="GI:2337975"  
/translation="MAHTGSTVCAFLIFAVLKNVFCOPTSSSEVEDVDEPANTVSDN  
IIHQQRNRTKGIHSDPAPFVGCASNIGDIFRQTSHCNPTDOKENESILLIF  
KENIVAVFVFKRYKRTIVTSTYNGIYDAVAVNOHVESVPIYFRBMDITYOCN  
SLDVTYGNGLIYTDNDGSMYTDIDLPVGLSLSVRYRHSQPIAHAPGMLLGGYRR  
TIVNCEVETIDARAAPPFRIFTINIGDTIEKSPFSANMETEFGSDPTLTVAKDY  
RVNDYKRGQPOGQHRIFDKBEYLLSMAQQRNLSYCMWAMKSFDAIKTEHKS  
LHVANDITAFSPYPTQTEVGLKHYCNLLTESELKSLANVNDHSNGADQYLL  
TNGGLLVQOPLQOOLKLDKGLADVAKQONTTITTTTTSRORSRVSAGIDVYT  
AESTIITLOTPAYDILRQIINNVELEISAMKROHRAKSIAMNELSKINPTSVMSI  
YGRPVSAKRGDIVISYSHCVVDQDSYSLRSKRVPRQDKTHCYSRPPTFFRIND  
HLKRGDGVNNEILLITTTVAETICHENHTEFOCGNMWYFKNRHAKTMAVGVATID  
TFVNLNLTIVENIDFOYIELYSREBKMSAFPIETEMFRYNNYTRQVGLRMDLFG  
ATNRNDFVDAGSLMDIDGVYATLVNASSVATLSSISYGIILNIFNGGMLLFG  
LIAAVYTVILRKRAKRFQONQVOMIYPIKITTOSRELOLYDPSIKHBLDITLAM  
HDVHASQPSRKQDEDEGSTTSQPDMLNKKANVLRRAQYKRLKRTDSESGVPE"  
19314..19319  
/evidence=not\_experimental  
19428..22508

polyA\_signal  
CDS















CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue as molecular weight markers and as  
CC a food supplement (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (II) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 693 BP: 296 A; 79 C; 234 G; 84 T; 0 other;

Alignment\_scores:  
Quality: 156.00 Length: 85  
Ratio: 2.438 Gaps: 2  
Percent Similarity: 75.294 Percent Identity: 41.176

alignment\_block:  
US-08-973-363-11 x AAS74240 ..

Align seg 1/1 to: AAS74240 from: 1 to: 693

```
11 LysLysLLeLysThrGluLysGluAsnGluLysProGluProAsp11 27
   ||||| ||| ||||| ||||| ||||| ||| |||
190 AAGAAAGATTAAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAA 237
27 eGlyLLeLysGluLysGluLysGluLysGluLysGluLysGlu 44
   ::||| ||||| ||||| ||||| ||||| |||||
238 .AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 286
44 snLysArgLysLeuLysArgLysGluLysGluLysGluLysGlu 60
   ::||| ||||| ||||| ||||| ||||| |||||
287 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 336
61 LeuLys.....GluLysAspAsnLysGluLysArg 70
   ||||| ||||| ||||| ||||| ||||| |||||
337 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 386
70 gGluAsnLysValLysGluSerThrGlnLysGluLysGluValLysGlu 87
   ::||| ||||| ||||| ||||| ||||| |||||
387 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 436
87 IuLys 88
   ::|||
437 AGAAG 441
```

seq\_name: /STDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAS90715

seq\_documentation\_block:  
ID AAS90715 standard; cDNA: 693 BP.

```
XX AAS90715:
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #26519.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX MO200175067-A2.
XX
XX 11-OCT-2001.
XX
```

PF 30-MAR-2001; 2001MO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
XX P-PDB: AAG26528.  
XX

XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX

XX Claim 1; SEQ ID NO 26519; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 693 BP: 296 A; 79 C; 234 G; 84 T; 0 other;

alignment\_scores:  
Quality: 156.00 Length: 85  
Ratio: 2.438 Gaps: 2  
Percent Similarity: 75.294 Percent Identity: 41.176

alignment\_block:  
US-08-973-363-11 x AAS90715 ..

Align seg 1/1 to: AAS90715 from: 1 to: 693

```
11 LysLysLLeLysThrGluLysGluAsnGluLysProGluProAsp11 27
   ||||| ||| ||||| ||||| ||||| ||| |||
190 AAGAAAGATTAAGAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAA 237
27 eGlyLLeLysGluLysGluLysGluLysGluLysGluLysGlu 44
   ::||| ||||| ||||| ||||| ||||| |||||
238 .AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 286
44 snLysArgLysLeuLysArgLysGluLysGluLysGluLysGlu 60
   ::||| ||||| ||||| ||||| ||||| |||||
287 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 336
61 LeuLys.....GluLysAspAsnLysGluLysArg 70
   ||||| ||||| ||||| ||||| ||||| |||||
337 AAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 386
70 gGluAsnLysValLysGluSerThrGlnLysGluLysGluValLysGlu 87
   ::||| ||||| ||||| ||||| ||||| |||||
387 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 436
```













Mon Aug 5 11:51:39 2002

us-08-973-363-11.p2n.rng

Page 11



**THIS PAGE BLANK (USPTO)**



[illegible]



[illegible]

Plate: LCM1033 row: k column: 18  
High quality sequence stop: 562.

# FEATURES

Location/Qualifiers  
1..821

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:413129"  
/clone\_lib="NH\_MGC-54"  
/issue\_type="from chronic myelogenous leukemia"  
/lab\_host="DH10B (T1 phage-resilient)"  
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech); Site\_1: Still (ggcgcctcgcgc); Site\_2: Still (ggccatctggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATATGACC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGCGGCGGCACATG-3' (30)BN-3' (where B = A, C or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT  
ORIGIN

345 a 149 c 190 g 137 t

## alignment\_scores:

Quality: 266.00 Length: 90  
Ratio: 3.367 Gaps: 2  
Percent Similarity: 87.778 Percent Identity: 68.889

## alignment\_block:

US-08-973-363-11 x BF239967 ..

Align seg 1/1 to: BF239967 from: 1 to: 821

```

1  ASGGLIIEVAlSeRVAllySHISleuHISlySLySIIleYstHrGluLy 17
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
234 GATGAGATCAGTCTGTGTAACATCAATATAAAAAATTTAAAAACAGAAAG 283
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
17  SGLU.ASnglUGLUySPrgLUpProASpIIleGlyIleYstHrGluA 33
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
284 AGACCAAGTGAAGAAAACTGTGAGCCAGATGTTTATATTAAGAAAGACCA 333
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
34  GLUglUySArgGLUthrLySGLUySGLUySGLUySArgGLUleUlySAr 50
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
334 GAAGAAAAAGAGAGCAAAAGCAAAAGAGATTAATAAAAGAACTTAAGAA 383
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
50  g.GLUyLySGLUySGLUySGLUySGLUySGLUySGLUySGLUySAspAsn 66
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
384 GTGAGATTAAGAAAAAGAGATTAAGAAAGATAT.AAGGACACAGATTTT 432
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
67  LysGLUySArgGLUySGLUySGLUySGLUySGLUySGLUySGLUySGL 83
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
433 AAAGAAAAAGAGAAACCAAAAGTAAAGAACTTATACAGAAAGAAAG 482
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
83  uVallySGLUyS 88
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
483 ACTAAAGACGAAAG 498

```

seq\_name: gb\_est2:BM168938

## seq\_documentation\_block:

LOCUS BM168938 660 bp mRNA linear EST 04-DEC-2001  
DEFINITION EST571461 PYBS Plasmodium yoelii yoelii cDNA clone pYCPJ56 5' end,  
mRNA sequence.

ACCESSION BM168938 GI:17302170  
VERSION BM168938.1

KEYWORDS EST.  
SOURCE Plasmodium yoelii yoelii.

# ORGANISM

Eukaryote: Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
REFERENCE 1 (bases 1 to 660)  
AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B.,

TITLE Fraser,C.M. and Carucci,D.J.  
JOURNAL Plasmodium yoelii EST project at TIGR  
COMMENT Unpublished (2001)

Contact: Jane Carlton  
Parasite Genomics Group  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-530-9319  
Fax: 301-838-0208  
Email: carlone@igf.org  
For clone info, please contact the Malaria Research and Reference  
Reagent Resource Center, ATCC  
http://www.malaria.mri.org/mr4pages/index.html  
Seq primer: ADP.

# FEATURES

Location/Qualifiers  
1..660

/organism="Plasmodium yoelii yoelii"  
/strain="17XL"  
/db\_xref="taxon:73239"  
/clone="PYCPJ56"  
/clone\_lib="PYBS"  
/dev\_stage="asexual blood stages"  
/lab\_host="E. coli XL-1 Blue"

/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cBYJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size fractionated cDNA was precipitated and ligated to HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

BASE COUNT  
ORIGIN

367 a 35 c 143 g 115 t

## alignment\_scores:

Quality: 195.00 Length: 78  
Ratio: 3.047 Gaps: 0  
Percent Similarity: 82.051 Percent Identity: 48.718

## alignment\_block:

US-08-973-363-11 x BM168938 ..

Align seg 1/1 to: BM168938 from: 1 to: 660

```

11  LysLySIIleYstHrGluYSLUySGLUySGLUySGLUySGLUySGLUyS 27
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
4  CCCGAGTTGAAGAACTGAGAAAGCAAAAGCAAAAGCAAAAGCAAA 53
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
27  eGLyIleYSLySGLUAlaGLUySArgGLUthrLySGLUySGLUySGLU 44
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
54  AGAAGAAAAAGAGATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 103
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
44  snLySArgGLUleUlySArgGLUySLySGLUySGLUySGLUySGLUyS 60
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
104 AAAAGAAAGATTAAGAGATTAAGAAAGAAAGAAAGAGATTAAGAA 153
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
61  LeUlySGLUySAspAsnLySGLUySArgGLUySGLUySGLUySGLUyS 77
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
154 GAAAGGAAGAGATTAAGAAAGAAAGAAAGAAAGATTAAGAGAGAGAA 203
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
77  rThrGlnLySGLUySGLUVallySGLUyS 88
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
204 AAAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAG 237

```

```
seq_name: qb_est2:BM161742
seq_documentation_block:
LOCUS      BM161742                605 bp    mRNA          linear   EST_04-DEC-2001
DEFINITION ESR564265 PYBS Plasmodium yoelii yoelii cDNA clone pYCK149 5' end,
            mRNA sequence.
ACCESSION  BM161742
VERSION    BM161742.1 GI:17307423
KEYWORDS   EST
SOURCE     Plasmodium yoelii yoelii.
ORGANISM   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 605)
AUTHORS    Carlson,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdy,A.B.,
            Fraser,C.M. and Carucci,D.J.
TITLE      Plasmodium yoelii EST project at TIGR
JOURNAL    Unpublished (2001)
COMMENT    Contact: Jane Carlton
            Parasite Genomics Group
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel.: 301-530-9319
            Fax: 301-838-0208
            Email: carlton@tigr.org
            For clone info, please contact the Malaria Research and Reference
            Reagent Resource Center, ATCC
            http://www.malaria.mri.org/mr4pages/index.html
            Seq primer: ADF.

FEATURES             Location/Qualifiers
     source           1..605
                     /organism="Plasmodium yoelii yoelii"
                     /strain="17XL"
                     /db_xref="taxon:73239"
                     /clone="PYCK149"
                     /cclone="PYCOP51"
                     /cclone.lib="PYBS"
                     /dev_stage="Asexual blood stages"
                     /lab_host="E. coli XL-1 Blue"
                     /note="Vector: PAD-GAL4. At 20-25% parasitemia, blood was
collected from BALB/cBY mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephadryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (PAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

BASE COUNT      342 a       37 c       138 g       88 t
ORIGIN
...
Alignment_scores:
Quality:      193.00      Length:      77
Ratio:        3.063      Gaps:      0
Percent Similarity: 81.818 Percent Identity: 49.351

alignment_block:
US-08-973-363-11 x BM161742 ..
Align seg 1/1 to: BM161742 from: 1 to: 605

12 LysatlelygthgtulysgluasngclutylspTCGCUPTroaspillect 28
8 GAAGTGAAATCGAGAAAGTGAAGAAGAAAGAAAGAAAAGAAAAAGAA 57
```

```
FEATURES      Location/Qualifiers
source        1. .507
```

seq\_name: gb\_est1:AI890775

seq\_documentation\_block:

LOCUS AI890775 547 bp mRNA linear EST 07-MAR-2000

DEFINITION um95f11.x1 NCI\_CGAP ut2 Homo sapiens cDNA clone IMAGE:2443725 3' similar to SW:CHD1.HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING PROTEIN 1 ; mRNA sequence.

ACCESSION AI890775

VERSION AI890775.1 GI:5595939

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

TITLE Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@db-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.

[illegible]





## COMMENT

This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

## FEATURES

## source

1. .981  
/organism="Debaryomyces hansenii"  
/strain="CBS 767"  
/variety="hansenii"  
/db\_xref="taxon:4959"  
/clone="BC0AA009H03"  
/clone\_1lb="BC0AA"  
/note="end : T3"  
<211. .>665  
/note="similar to Saccharomyces cerevisiae ORF YGR280C [  
weak similarity to Cbf5p ]  
1 putative frameshift(s)"  
/evidence="not\_experimental"

## misc\_feature

/note="similar to Saccharomyces cerevisiae ORF YGR280C [  
weak similarity to Cbf5p ]  
1 putative frameshift(s)"  
/evidence="not\_experimental"

BASE COUNT 430 a 98 c 241 g 209 t 3 others  
ORIGIN

## alignment\_scores:

Quality: 170.00 Length: 94  
Ratio: 2.656 Gaps: 4  
Percent Similarity: 68.085 Percent Identity: 45.745

## alignment\_block:

US-08-973-363-11 x CNS07B7L ..

Align seg 1/1 to: CNS07B7L from: 1 to: 981

```
5 SerValTysHisLeuHisLysLysIle..... 13
||| ::::||| |||||
711 AGTAGCAGCATTAAGACAAAAATAAAGAGACCGTAAGAAAAAGA 760
14 .....LysThrGluLysGluAsnGluGluLysProGluProAspIleG 28
||||| ||||| ||||| ||||| |||||
761 AGAGAGAGAGACCGAGAGAAATATGTGAG..... 791
28 LysLeuLysLysGluLysGluLysArgGluThrLysGluLysGluAsn 44
||||| ||||| :::: ||||| ||||| |||||
792 ..ATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 839
45 LysArgGluLeuLysArgGluLysLysGluLysGluAspLysLysGluLe 61
||| :::: ||||| ||||| ||||| ||||| |||||
840 AAGARAGAT...AAGACGAGAGAGAGAGAGAGAGATTAAGACAGAGA 886
61 ULysGluLysAspAsnLysGluLysArgGluAsnLysValLysGluSert 78
||||| ||||| ||||| ||||| ||||| |||||
887 GAAAGAGAGAG.....AAGATTAAGACGAGAGAGAGAGAGAGAGAGA 930
78 hrGluLysGluLysGluValLysGluLysGluLys 88
::::: ||||| ||||| ||||| ||||| |||||
931 AGAAGAGAGAGAGAGAGAGAGATTAAGAGAA 962
```





MEDLINE 97473516  
REFERENCE 2 (bases 1 to 6872)  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK  
FEATURES  
source Location/Qualifiers  
gene /organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1..6872  
/gene="CHD-2"  
228..5654  
/gene="CHD-2"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the 2 chromosome  
protein"  
/protein\_id="AAC60282.1"  
/db\_xref="GI:2501846"  
/translation="MNGHSDSESVRNSGGSSRDDSSAGSGSGSSSSSSDSS  
SOGSSDSESGSSGSSSESDTSREKKYQAKPPRADSEKSSPSLAVQRAV  
LKKQQQQAASDSSGSESDSSADSSSETKKKKRDWQSSGVSGETS  
DSSADGDKSSCESESDYEPKVKSRPSPRIKSGKSGTKGKRLDSEED  
DDDEDKRGSRQATVNSYKAEAEETKTDSDDLCEYCGDVEQTEDEFEETEKMD  
SRIRKATGASTTIAVEADGDPNAGFEKSELGEIOYLKKMGSHINTEET  
LKOQVKGNNKLDNYKKKQDETQRLKNAPEVEYXNCOQLTDLQKYOYERI  
AHNOKSAAGYPDYKMGQIPYSECSMEGALAKFKQRIEYPSRNSKTPED  
CKYKORPRVALKOPSTIGHESLELDYQLNGLWLHSMCKSGKSLADEMGIG  
KITQTSFLNLYLHERQLYGPFLRLVPLSTLWQREIQTMAQOMNAVVLGDTISN  
MITHEMNHPOTKRLKFNILLTYEILKDKSLFGVDEAHRLKNDLSLY  
RLIDKSNHRLITGTPONSLKELMSLHFMPEKFSMEDEEHGGREYGS  
LHKELEPELLRRVKDKVEKSLPAKVEQILMEKSALOKYYKMLITRYKALSKSGK  
STSGFLINMELKCCNHCYLIKRPDNNYNNQALQHLIRSGKLIIDKLIR  
ERGNRVLPESQWMDITAEIKYRPPORDGSRGELKQALDHFAESSEDC  
FLSTRAGGLINLASADYVIFDSMNPNQDLOQARAHRIQKQKVNIYRLVTGS  
VEEDILERAKKVVLDHLVQRMDTGKYLHSGSPSSPTPNKELSLILFGAE  
LFKEPGEQEPQEMDIDELIKRAETRENEPGLYVDELISOFYANFEMDEDE  
LEPERNSRMEEIIPESORRIIEEEROKLEIETVLPFRMNCAGKISFGSGRRR  
SRYSGSDSDSITERKPKRGRPRTPRENIIGFSAEIRRPITKSKKRGGLERD  
AVARDAELDKSETDLRLGELYHNGCIKRLAKNSSQGEAAGRLKVKPPTRIISY  
QVNAKLYISHEELAPLHKSIPSDPERKRYVPCHTKAHFIIDWKEKDSNLVGI  
YEGYGSWEMIAMDPLSLQKILPPDPDKPAKQIADVLKILKNDLARKKEQ  
RLAGAGNSKRRTKRNKMKAKSIKEEISDSSDPSSESDDEEDKDDIVSYK  
HLHKIKITEKENEKEPDDIGIKKEAEKRETKEREKRLKREKREKKEKLEKED  
NKKRENNKYESTQKEKEVEKVENKSKKSKKIPLDTPVHTITASEPVPISE  
ESBELHOKTESVCKEEMRPYKALKQDREPKEKLSEREOLEHTRQCIKIGDHITBEL  
KETNPBQIKQKRNIMIFYSKTFEDARKLKLKALKQESQOQHNQNTSSNV  
THVIRNPDYERLKETTNHDSRDSYSDRHLIQYHDHNRQGAAYKRSDSKRPY  
SAFNSGKRDHMDWLYKODSRYSDSKHKRLDHRSDHRSNLEGNLKDSSGHDHRS  
SDHRIHSDHRSSTSEYSHKSKSDYRYHSDWQMDHRAASGSPRLDQSPYSGRSPFG  
HRSPEHSSDHKSTPEHTWSSRKT"  
4341..4604  
/gene="CHD-2"  
/note="short insert found in longer variant mRNA of CHD-2"  
BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN  
misc\_feature  
4341..4604  
/gene="CHD-2"  
/note="short insert found in longer variant mRNA of CHD-2"  
BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN  
alignment\_scores:  
Quality: 193.00 Length: 45  
Ratio: 4.707 Gaps: 0  
Percent Similarity: 91.111 Percent Identity: 86.667  
alignment\_block:  
US-08-973-363-14 x AF004397 ..  
Align seg 1/1 to: AF004397 from: 1 to: 6872  
1 AspglyIleValSerValIshisProHstLysLysIleLysAlaGlu 17  
|||||  
4341 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAGCAAAAA 4390

17 sGLuAsnGLuGLuAspgluProGluIleGlyIleLysLysGluAlaG 34  
|||||  
4391 AGAAAATGAGAAAAAGCCGACCGACGATRTTGTCTAAAGAGAACTG 4440  
34 IyGLuLysArgGluThrLysGluLysGluAsnLys 45  
|||||  
4441 AAGAAAAAGAGACAGACAAAAGAGAAAGAAATATAA 4475  
seq\_name: gb\_pat:A58693  
seq\_documentation\_block:  
LOCUS A58693 265 bp DNA linear PAR 06-MAR-1998  
DEFINITION Sequence 12 from Patent WO9639505.  
ACCESSION A58693  
VERSION A58693.1 GI:3714251  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 265)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS  
JOURNAL Patent: WO 9639505-A 12 12-DEC-1996;  
COMMENTS ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
source Location/Qualifiers  
1..265  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 158 a 16 c 61 g 30 t  
ORIGIN  
alignment\_scores:  
Quality: 180.00 Length: 46  
Ratio: 4.390 Gaps: 1  
Percent Similarity: 89.130 Percent Identity: 84.783  
alignment\_block:  
US-08-973-363-14 x A58693 ..  
Align seg 1/1 to: A58693 from: 1 to: 265  
1 AspglyIleValSerValIshisProHstLysLysIleLysAlaGlu 17  
|||||  
1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATAAAGCAAAAA 50  
17 ysgLuAsnGLuGLuLysAspgluProGluIleGlyIleLysLysGluAla 33  
|||||  
51 AAGAAATGAGAAAAAGCCTGAGCCAGATATGCTATAAAGAGAACT 100  
34 GlyLysArgGluThrLysGluLysGluAsnLys 45  
|||||  
101 GAAAGAAAAAGAGACAAAAGAGAAAGAAATATAA 136  
seq\_name: gb\_pr:AC092372  
seq\_documentation\_block:  
LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001  
DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
ACCESSION AC092372  
VERSION AC092372.3 GI:17402768  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 101220)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished

```

REFERENCE 2 (bases 1 to 101220)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 101220)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
JOURNAL Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 7, 2001 this sequence version replaced gi:15290448.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 146,7kb). It is clipped at the overlap with AC012624.
The number of bases overlapped is 90404.
FEATURES
source
1..101220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-58M12"
BASE COUNT 34122 a 18862 c 17827 g 30409 t
ORIGIN
alignment_scores:
Quality: 159.00 Length: 45
Ratio: 4.077 Gaps: 0
Percent Similarity: 86.667 Percent Identity: 68.889
alignment_block:
US-08-973-363-14 x AC092372/rev ..
Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220
1 AsgclYlIeValserVallyshisProhislyslYslYslalGcluly 17
|||||
25410 GATGAGATCACTTCGTGAAACATCCAAATTAACCAAGAAAG 25361
17 sgluYsnglucglulYsAspgcluprogluIleaglylYslYslgluYslag 34
|||||
25360 AGCAGGTGAAAGAAAACCTGAGCCAGATGTTATATTAAGAGAAACGCG 25311
34 lyglulYsArGclurthrlYsGluYslgluYsluAsnly 45
|||||
25310 AAGAAAAGAGGAAAGCAAGAAAAGAGAGATTAAG 25276
seq_name: gb_pr:AC012624
seq_documentation_block:
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-208211, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:14993679
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 134365)
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
REFERENCE 2 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
SUBMITTED (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

```

[illegible]









*	27984	30366: contig of 2403 bp in length
*	30387	30466: gap of unknown length
*	30487	32480: contig of 1994 bp in length
*	32481	32550: gap of unknown length
*	32581	35441: contig of 2861 bp in length
*	35442	35541: gap of unknown length
*	35542	37800: contig of 2259 bp in length
*	37801	37900: gap of unknown length
*	37901	40587: contig of 2687 bp in length
*	40588	40687: gap of unknown length
*	40688	44159: contig of 3472 bp in length
*	44160	44259: gap of unknown length
*	44260	46636: contig of 2377 bp in length
*	46637	46736: gap of unknown length
*	46737	50082: contig of 3346 bp in length
*	50083	50182: gap of unknown length
*	50183	53988: contig of 3806 bp in length
*	53989	54088: gap of unknown length
*	54089	56592: contig of 2504 bp in length
*	56593	56692: gap of unknown length
*	56693	61332: contig of 4660 bp in length
*	61353	61432: gap of unknown length
*	61453	64234: contig of 2802 bp in length
*	64255	64354: gap of unknown length
*	64355	68825: contig of 4471 bp in length
*	68826	68925: gap of unknown length
*	68926	74393: contig of 5468 bp in length
*	74394	74493: gap of unknown length
*	74494	80551: contig of 6068 bp in length
*	80562	80661: gap of unknown length
*	80662	87656: contig of 6965 bp in length
*	87627	87726: gap of unknown length
*	87727	93600: contig of 5874 bp in length
*	93601	93700: gap of unknown length
*	93701	102024: contig of 8324 bp in length
*	102025	102124: gap of unknown length
*	102125	109905: contig of 7781 bp in length
*	109906	110005: gap of unknown length
*	110006	116737: contig of 6732 bp in length
*	116738	116837: gap of unknown length
*	116838	126797: contig of 9960 bp in length
*	126798	126897: gap of unknown length
*	126898	135114: contig of 8217 bp in length
*	135115	135214: gap of unknown length
*	135215	146713: contig of 11499 bp in length
*	146714	146813: gap of unknown length
*	146814	157894: contig of 11081 bp in length
*	157895	157994: gap of unknown length
*	157995	171752: contig of 13758 bp in length
*	171753	171852: gap of unknown length
*	171853	201867: contig of 29835 bp in length
*	201868	201787: gap of unknown length
*	201788	231268: contig of 29481 bp in length
*	231269	231368: gap of unknown length
*	231369	276181: contig of 44813 bp in length

FEATURES	location/Qualifiers
1. .276181	
source	
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="5"
	/clone="RP11-75H1"
	/clone_id="RPCI human BAC library 11"
BASE COUNT	86859 a 51769 c 50026 g 82854 t 4673 others
ORIGIN	

```
alignment_scores:          Length:    45  
      Quality:        159.00  
      Ratio:         4.077  
Percent Similarity:     86.667   Percent Identity: 68.889
```

alignment\_block:  
US-08-973-363-14 x AC092382/rev .

```

Align seg 1/1 to reverse of: AC092382 from: 1 to: 276181

1 aspgylilevalservallyshsprofhslysllyllylalyalaguly 17
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
212172 gattgagatagctctgtgaaacatccaaattaaaaattaaaaacgnaag 212123
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 sgluansngluglulyaspgluproglutlleglyllyllylsglualag 34
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
212122 aacacagtgaaagaaaaccctgagccacagatctttatattaaaagaaagacacag 212073
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 lyglulysarngluthrlysglulysgluasnlly 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
212072 aagaaaaagcgagacgaaagaaaaggaatataa 212038
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: gb_pr:HS1185H19

```

seq\_documentation\_block:  
LOCUS HS1185H19 104780 bp DNA linear PRI 27-SEP-2000  
DEFINITION Human DNA sequence from clone Rp5-1185H19 on chromosome 1p13.1-13.3  
Contains part of a gene for a novel protein, STSS, GSSS and a Cpg island, complete sequence.  
ACCESSION AL121982  
VERSION AL121982.19 GI:9187337  
KEYWORDS HTG; CpG Island.  
SOURCE human.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Eukaryota: Chordata: Craniala: Vertebrata: Euteleostomi:  
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.  
1 (bases 1 to 104780)  
Cobley, V.  
Direct Submission  
Submitted (21-AUG-2000) Sanger Centre, Hinxton, Cambridgeshire,

COMMENT On Jul 14, 2000 this sequence version replaced gi:8919096.

COMMENT

On Jul 14, 2000 this sequence version replaced g1:8919096.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/RCF/Chrl>

RP5-1185H19 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pCYPAC2

This sequence is the entire insert of clone RP5-1185H19 The true right end of clone RP5-1039JL1 is at 65940 in this sequence.

FEATURES	Location/Qualifiers
SOURCE	1. .104780

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p13.1-13.3"
/clone="RP5-1185H19"
/clone_1b="RPCL-5"
26.677
repeat_region
/note="11M4 repeat: matches 2452.3097 of consensus

```







```

repeat_region /rpt_family="AT_r1ch"
complement(1491..1637)
/rpt_family="B1_M1"
repeat_region 1648..1688
/rpt_family="AT_r1ch"
repeat_region 1786..1822
/rpt_family="(TTTC)n"
complement(1828..1958)
/rpt_family="B1_M1"
repeat_region complement(2006..2367)
/rpt_family="ORR1C"
repeat_region 3600..3772
/rpt_family="B3"
repeat_region 3773..3833
/rpt_family="(CA)n"
repeat_region 3834..3850
/rpt_family="B3"
complement(4270..4469)
/rpt_family="B3A"
repeat_region 5341..5370
/rpt_family="AT_r1ch"
repeat_region 5331..5772
/rpt_family="MTE"
repeat_region 5774..5803
/rpt_family="(TTTG)n"
complement(5829..5845)
/rpt_family="B2_M2"
repeat_region 5846..5926
/rpt_family="(TA)n"
complement(5927..6073)
/rpt_family="B2_M2"
repeat_region 6078..6138
/rpt_family="MTE"
repeat_region 6443..6622
/rpt_family="L1M1"
complement(7600..8061)
/rpt_family="B1_M1"
repeat_region complement(7603..7618)
/rpt_family="B4A"
complement(7619..7931)
/rpt_family="ORR13"
complement(7932..8116)
/rpt_family="B4A"
repeat_region 8518..8883
/rpt_family="(CA)n"
repeat_region 9316..9337
/rpt_family="AT_r1ch"
complement(9373..9534)
/rpt_family="KSTINE1"
repeat_region 9690..9709
/rpt_family="(CAA)n"
repeat_region 9932..9978
/rpt_family="(CA)n"
repeat_region 10152..10251
/rpt_family="(CA)n"
repeat_region 10857..10882
/rpt_family="(CA)n"
complement(10882..10882)
/standard_name="low coverage"
note="GGA Repeat"
repeat_region 11222..11273
/rpt_family="(GCA)n"
repeat_region 11600..11629
/rpt_family="(G)n"
complement(11658..11712)
/rpt_family="P81D7"
repeat_region 13772..13944
/rpt_family="RMR17C"
repeat_region 13945..13993
/rpt_family="(TC)n"
repeat_region 13994..14145
/rpt_family="RMR17C"
complement(14702..15092)
/rpt_family="M7B"
repeat_region complement(15182..15485)

```

```

alignment_scores:
  Quality: 87.50 Length: 45
  Ratio: 2.734 Gaps: 1
  Percent Similarity: 71.111 Percent Identity: 44.444

alignment_block:
US-08-973-363-14 x AC067964 ..

Align seg 1/1 to: AC067964 from: 1 to: 219769

1 AspglylEvaIServallYsHsProhlsYsLyslElysaAlglulY 17
|||||.....:|||||
201985 GATTGGGAATACAGTTCAGTATGACTGGAAAAA.....GAAGA 202025
17 sgluasnglulysaspluprogulleglylElysgluAag 34
|||||.....:|||||
202026 TGAAGAAGAGGAGAGAGAGAGAGAGAGGTGAAGAAGAGAAAG 202075
34 lyglulysarglulThlysglulysgluAsnlys 45
|||||.....:|||||
202076 GAGAGAGAGAGGAGAGAGAGAGAGAGAGG 202110
seq_name: gb_ro:AL589701

```

```

seq_documentation_block:
LOCUS AL589701 219200 bp DNA linear ROD 30-JAN-2002
DEFINITION Mouse DNA sequence from clone RP23-202F3 on chromosome 13.
ACCESSION AL589701
VERSION AL589701.9 GI:18476660
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathu; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 219200)
AUTHORS Phillimore, R.
TITILE Direct Submission
JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Feb 1, 2002 this sequence version replaced gi:17384104.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL; Sw.,
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-202F3 is
from the RP23-23 Mouse PAC library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: PBAC3.6
This sequence is the entire insert of clone RP23-202F3. The true
left end of clone RP23-171015 is at 179956 in this sequence. The
true right end of clone RP23-15386 is at 117857 in this sequence.

```

```

COMMENT

```

```

FEATURES
  source
    1..219200
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /chromosome="13"
      /clone_id="RP23-202F3"
      /clone_11b="RPCT-23"
      3942..4127
        /note="Single clone region. Sequence from clone PCR only."
      29783..30099
        /note="Single clone region. Sequence from clone PCR only."
      misc_feature
        /note="Single clone region. Sequence from clone PCR only."
      misc_feature
        /note="Single clone region. Sequence from clone PCR only."
      short insert library derived from a clone PCR. Restriction
      digest data confirm the assembly."
BASE COUNT
  63635 a 46933 c 46966 g 61646 t
ORIGIN

alignment_scores:
  Quality: 87.00 Length: 35
  Ratio: 3.346 Gaps: 0
  Percent Similarity: 74.286 Percent Identity: 57.143

alignment_block:
  US-08-973-363-14 x AL589701/rev ..

Align seg 1/1 to reverse of: AL589701 from: 1 to: 219200

11 LysylsileylsAlaGluysAluasnGluysAspGluProGluI 27
||||| ||| |||||||:|||||:|||||:|||||
61324 AAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61275
27 eGlyIleLysGluAlaGluysArgGluThrLysGluysGluA 44
|||||:||||| |||||||:|||||:|||||:|||||
61274 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 61225
44 snlys 45
:::||||
61224 AGAAG 61220

seq_name: gb_htg:AC098600

seq documentation block:
LOCUS AC098600 231947 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus chromosome R1 clone CH230-95C8, WORKING DRAFT
SEQUENCE 62 unordered pieces.
ACCESSION AC098600
VERSION AC098600.4 GI:17973757
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 231947)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-oshman,F.R., Allen,C.,
Alsprouck,S.L., Amaratunge,H.C., Aye,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burke,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Dem,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Roches,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Fritsch,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Gara,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,M., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homsl,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT

2 (bases 1 to 231947)
Unpublished
Direct Submission
Worley,K.C.
Submitted (26-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064593.

----- Genome Center of Medicine
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GIKD
Center clone name: CH230-95C8
----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findhaplolist
Consensus quality: 191710 bases at least Q40
Consensus quality: 199153 bases at least Q30
Consensus quality: 203798 bases at least Q20
Estimated insert size: 197288; sum-of-ctrls estimation
Quality coverage: 0x in Q20 bases; agarose-IP estimation
Quality coverage: 3.1x in Q20 bases; sum-of-ctrls estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
11490: contig of 11490 bp in length
11491 11590: gap of unknown length
11591 21783: contig of 10193 bp in length
21784 21883: gap of unknown length
21884 30102: contig of 8219 bp in length
30103 30202: gap of unknown length
30203 40486: contig of 10284 bp in length
40487 40587: gap of unknown length
40588 47306: contig of 6720 bp in length
47307 47406: gap of unknown length
47407 56516: contig of 9110 bp in length
56517 56617: gap of unknown length
56618 64419: contig of 7803 bp in length
64420 64519: gap of unknown length
64520 71196: contig of 6677 bp in length
71197 71197: gap of unknown length

```

```
* 71297 77926: contig of 6630 bp in length
* 77927 77927: gap of unknown length
* 78027 84372: contig of 6346 bp in length
* 84373 84373: gap of unknown length
* 84473 90426: contig of 5957 bp in length
* 90430 90530: gap of unknown length
* 90530 94563: gap of unknown length
* 94563 94570: gap of unknown length
* 94570 100480: contig of 3811 bp in length
* 100480 100580: gap of unknown length
* 100580 100586: contig of 5056 bp in length
* 100586 105737: gap of unknown length
* 105737 105736: contig of 4106 bp in length
* 105736 109842: gap of unknown length
* 109842 109943: gap of unknown length
* 109943 114558: contig of 4916 bp in length
* 114558 114559: gap of unknown length
* 114559 118534: contig of 3576 bp in length
* 118534 118535: gap of unknown length
* 118535 123663: contig of 5029 bp in length
* 123663 123763: gap of unknown length
* 123763 127449: contig of 3985 bp in length
* 127449 12748: gap of unknown length
* 12748 131812: contig of 3964 bp in length
* 131812 131912: gap of unknown length
* 131912 135499: contig of 3587 bp in length
* 135499 135500: gap of unknown length
* 135500 139874: contig of 4275 bp in length
* 139874 139974: gap of unknown length
* 139974 144298: contig of 4324 bp in length
* 144298 144398: gap of unknown length
* 144398 148526: contig of 4128 bp in length
* 148526 148626: gap of unknown length
* 148626 152730: contig of 4104 bp in length
* 152730 152830: gap of unknown length
* 152830 152831: contig of 4749 bp in length
* 152831 157579: gap of unknown length
* 157579 157679: gap of unknown length
* 157679 161266: contig of 3587 bp in length
* 161266 161366: gap of unknown length
* 161366 164887: gap of unknown length
* 164887 165887: contig of 3621 bp in length
* 165887 167888: gap of unknown length
* 167888 170166: contig of 2901 bp in length
* 170166 170266: gap of unknown length
* 170266 172944: contig of 2678 bp in length
* 172944 173044: gap of unknown length
* 173044 176120: contig of 3076 bp in length
* 176120 176220: gap of unknown length
* 176220 178990: contig of 2770 bp in length
* 178990 179090: gap of unknown length
* 179090 181701: contig of 2611 bp in length
* 181701 181801: gap of unknown length
* 181801 183811: contig of 2010 bp in length
* 183811 183911: gap of unknown length
* 183911 185953: contig of 2042 bp in length
* 185953 186053: gap of unknown length
* 186053 187891: contig of 1838 bp in length
* 187891 187991: gap of unknown length
* 187991 190795: contig of 2804 bp in length
* 190795 190895: gap of unknown length
* 190895 193028: contig of 2133 bp in length
* 193028 193128: gap of unknown length
* 193128 195041: contig of 1913 bp in length
* 195041 195141: gap of unknown length
* 195141 196416: contig of 1275 bp in length
* 196416 196516: gap of unknown length
* 196516 198846: contig of 2330 bp in length
* 198846 198946: gap of unknown length
* 198946 201855: contig of 2309 bp in length
* 201855 201955: gap of unknown length
* 201955 203450: contig of 1495 bp in length
* 203450 203550: gap of unknown length
* 203550 206062: contig of 2512 bp in length
```

```
* 206063 206162: gap of unknown length
* 206163 207721: contig of 1559 bp in length
* 207722 207821: gap of unknown length
* 207822 210574: contig of 2753 bp in length
* 210575 210675: gap of unknown length
* 210675 212188: contig of 1514 bp in length
* 212188 212289: gap of unknown length
* 212289 213701: contig of 1413 bp in length
* 213701 213802: gap of unknown length
* 213802 215477: contig of 1746 bp in length
* 215477 215648: gap of unknown length
* 215648 217053: contig of 1406 bp in length
* 217053 217153: gap of unknown length
* 217153 218235: contig of 1081 bp in length
* 218235 218334: gap of unknown length
* 218335 219342: contig of 1008 bp in length
* 219343 219442: gap of unknown length
* 219443 220550: contig of 1108 bp in length
* 220551 220650: gap of unknown length
* 220651 222211: contig of 1561 bp in length
* 222212 222312: gap of unknown length
* 222312 224192: contig of 1781 bp in length
* 224192 224192: gap of unknown length
```

alignment\_scores:                   Quality: 87.00                   Length: 43  
                                     Ratio: 2.636                   Gaps: 0  
Percent Similarity: 76.744           Percent Identity: 39.535

alignment\_block:

US-08-973-363-14 x AC098600/rev ..

Align seg 1/1 to reverse of: AC098600 from: 1 to: 231947

```
3 11e1a1SeVAllyHs1pRoH1sYs1leYs1leYs1laGlulYsGlus 19
21458 ATGTTAAATGTTAAACATACGAAAGAAAGAGAGAGAGAGAGAG 21409
19 ngluGlulYsAspGluProGluTlleGlyTleYsYsGlulAglYglul 36
21408 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21359
36 ysArGluThrYsGluYsGluAsnLYs 45
21358 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 21330
```

---



17 ysgluasnGlulLysaspGluProGluIleGlyIleLysLysGluAa 33  
 51 AAGAAATGAGAAAAAGATGAGCCAGATTGTGATTAAGAGAGACT 100  
 34 GlyGluysArgGluThrLysGluLysGluLys 45  
 101 GGAGAAAAAGAGACAAAGAGAGAAATTAAG 136  
 seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142752

seq\_documentation\_block:  
 ID AA142752 standard; cDNA; 265 BP.

AC AAT42752;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chicken CHD-1A insert motif.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
 KM CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_difference 52 /tag= a  
 FT /note= "base 52 disrupts the reading frame for  
 FT the translated amino acid sequence given  
 FT in fig 7"  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96MO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 DR P-PSDB; AAW08144.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Disclosure; Fig 7; 76pp; English.  
 XX  
 CC A composite nucleotide sequence (AA142752) and putative translation  
 CC (AAW08144) sequence are provided of a motif that is found spliced to  
 CC a proportion of chicken CHD-1A clones. The motif is inserted  
 CC between bases 4327 and 4328 of the CHD-1A composite sequence  
 CC (AA142751). None of the 7 CHD-1 clones examined contained the  
 CC complete motif. There are no splice donor or acceptor sites within  
 CC the motif suggesting it is a final rather than an intermediary  
 CC product of splicing. The motif is also found at the 3' end of the  
 CC CHD-W clone CC14 (see also AA142753).  
 XX  
 SQ Sequence 265 BP; 158 A; 16 C; 61 G; 30 T; 0 other;

alignment\_scores:  
 Quality: 180.00 Length: 46  
 Ratio: 4.390 Gaps: 1  
 Percent Similarity: 89.130 Percent Identity: 84.783

alignment\_block:  
 US-08-973-363-14 x AA142752 ..

Align seg 1/1 to: AA142752 from: 1 to: 265

1 AspGlyIleValSerValLysHisProHisLysLysLysAlaGlu.L 17  
 1 GATGAGATTGTTTCAGTGAACATCTACATATAAAATTAACAGAAAA 50  
 17 ysgluasnGlulLysaspGluProGluIleGlyIleLysLysGluAa 33  
 51 AAGAAATGAGAAAAAGCCTGAGCCAGATTGTGATTAAGAGAGACT 100  
 34 GlyGluysArgGluThrLysGluLysGluLys 45  
 101 GAAGAAAAAGAGACAAAGAGAGAAATTAAG 136  
 seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA64139

seq\_documentation\_block:  
 ID AAA64139 standard; DNA; 50000 BP.

AC AAA64139;  
 XX  
 DT 20-DEC-2000 (first entry)  
 XX  
 DE Nucleotide sequence of a beta-tubulin antigen.  
 XX  
 KW Beta-tubulin antigen; inner ear protein; Meniere's disease; autoantibody;  
 KM chronic ear disease; autoimmune disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200050593-A1.  
 XX  
 PD 31-AUG-2000.  
 XX  
 PF 25-FEB-2000; 2000MO-US04795.  
 XX  
 PR 25-FEB-1999; 99US-0121549.  
 XX  
 PA (UYTE-) UNIV TENNESSEE RES CORP.  
 XX  
 PI YOO TJ;  
 XX  
 DR WPI; 2000-558400/51.  
 XX  
 PT New beta-tubulin antigen in the membranous structure of the inner ear,  
 PT reactive with antibodies of patients with Meniere's disease, for  
 PT diagnosing Meniere's disease and distinguishing this disease from other  
 PT autoimmune ear diseases  
 XX  
 PS Claim 3; Page 51-74; 115pp; English.  
 XX  
 CC The present sequence encodes a beta-tubulin antigen. The protein is  
 CC an antigen of the membranous structure of the inner ear protein, and  
 CC is reactive with antibodies from patients having Meniere's disease.  
 CC Meniere's disease is a chronic ear disease with unknown etiology.  
 CC Serum from patients suffering from this disease contain autoantibodies  
 CC against a 30 kDa cochlear protein antigen. The disease is believed to be  
 CC an autoimmune disease. The beta-tubulin antigen is useful as a target  
 CC substance in diagnosing or detecting Meniere's disease and in  
 CC distinguishing this disease from other autoimmune ear diseases.  
 XX  
 SQ Sequence 50000 BP; 16814 A; 10211 C; 9526 G; 13449 T; 0 other;

alignment\_scores:  
 Quality: 82.00 Length: 39  
 Ratio: 3.154 Gaps: 1  
 Percent Similarity: 66.667 Percent Identity: 51.282

alignment\_block:  
 US-08-973-363-14 x AAA64139 ..

Align seg 1/1 to: AAA64139 from: 1 to: 50000





PA	(HUMA- ) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Barash SC, Ruben SM,
XX	
XX	WPI; 2001-483426/52.
DR	

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 XX  
 PS Disclosure; SEQ ID NO 25592; 3071bp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosstatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 37783 BP; 9418 A; 9546 C; 9545 G; 9274 T; 0 other;

alignment\_block:

Align seg 1/1 to reverse of: AAK70780 from: 1 to: 37783

```
7 LysHisProHISLysLysIleLysAlaGluLysGluAsnGluGluLysAs 23  
||||| ::||| |::| |||||::| |||||::|  
16327 AAGAAATAAACACAGACAGCAGAAGAGAAGAGAAGAGAAGAGAAGGA 16278
```

23 pglurproctgltleeglylerylsyysgluuataglyglutlysargltutht 40  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
16277 gcagcaaaacgcagcagcgacgaagaagcagcagcagcagcacagcagaca 16228  
  
40 ysglutlysglutasnlys 45  
||||| ||| :::::  
16227 agcagaaagcaaaagcag 16211

```
seq_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK76625
seq_documentation_block:
ID   AAK76625 standard; DNA; 37783 BP.
```

XX  
AC      AAK76625;

DT 07-NOV-2001 (first entry)

DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:31437

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
KW

OS Homo sapiens.

PN W0200157182-A2.

PD 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US01354.

PR 31-JAN-2000; 2000US-0179065.

PR 24-FEB-2000; 2000US-0184664.

PR 16-MAR-2000; 2000US-0189874.

PR 18-APR-2000; 2000US-0198123.

PR 07-JUN-2000; 2000US-0209467.

PR 30-JUN-2000; 2000US-0215135.  
07-TMT-2000; 2000TC-0215547

PR	07-JUL-2000;	2000US-02168880.
DD	11-SEP-2000;	2000US-03174887.

PR 11-JUL-2000; 2000US-0217496.  
14 THU 2000UTC 0218200

PR 26-JUL-2000; 2000US-0220963 -  
DE 26-JUL-2000; 2000US-0220964

PR 14-AUG-2000; 2000US-0224518 -  
14-AUG-2000; 2000US-0224518 -

PR 14-AUG-2000; 2000US-0225213.  
14-AUG-2000; 2000US-0225214

PR 14-AUG-2000; 2000US-0225266-  
14-AUG-2000; 2000US-0225267

PR 14-ADG-2000; 2000US-0225268.  
14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.  
14-AUG-2000; 2000US-0225777

PR 14-AUG-2000; 2000US-0225758.  
DE 14-AUG-2000; 2000US-0225758.

PR	18-AUG-2000;	2000US-0226279.
PR	22-AUG-2000;	2000US-0236581.

PR 22-AUG-2000; 2000US-0226868.  
33-AUG-2000; 2000US-0327183.

PR	23-AUG-2000; 2000US-022/009.
PR	30-AUG-2000; 2000US-0328934.

PR 01-SEP-2000; 2000US-0229287.  
PB 01-SEP-2000; 2000US-0329343

PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.

PR	05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0338513	

PR 06-SEP-2000; 2000US-0230437-

[illegible]

```
|||||.....
16227 AGGAGAGGAGGAGGAG 16211

seq_name: /SIDSJ/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK80913

seq_documentation_block:
ID   AAK80913 standard; DNA; 37783 BP.
AC   AAK80913;
XX
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:35725.
XX
XX
KW   Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
XX   cytostatic; gene therapy; vaccine; metastasis; ds.
OS   Homo sapiens.
XX
XX   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
XX   31-JAN-2000; 2000US-0179065.
XX   04-FEB-2000; 2000US-0180628.
XX   24-FEB-2000; 2000US-0184664.
XX   02-MAR-2000; 2000US-0186350.
XX   16-MAR-2000; 2000US-0189874.
XX   17-MAR-2000; 2000US-0190076.
XX   18-APR-2000; 2000US-0198123.
XX   19-MAY-2000; 2000US-0205515.
XX   07-JUN-2000; 2000US-0209467.
XX   28-JUN-2000; 2000US-0214886.
XX   30-JUN-2000; 2000US-0215135.
XX   07-JUL-2000; 2000US-0216647.
XX   07-JUL-2000; 2000US-0216880.
XX   11-JUL-2000; 2000US-0217487.
XX   14-JUL-2000; 2000US-0217496.
XX   26-JUL-2000; 2000US-0218290.
XX   26-JUL-2000; 2000US-0220963.
XX   14-AUG-2000; 2000US-0224518.
XX   14-AUG-2000; 2000US-0224519.
XX   14-AUG-2000; 2000US-0225213.
XX   14-AUG-2000; 2000US-0225214.
XX   14-AUG-2000; 2000US-0225266.
XX   14-AUG-2000; 2000US-0225267.
XX   14-AUG-2000; 2000US-0225268.
XX   14-AUG-2000; 2000US-0225270.
XX   14-AUG-2000; 2000US-0225447.
XX   14-AUG-2000; 2000US-0225757.
XX   14-AUG-2000; 2000US-0225758.
XX   14-AUG-2000; 2000US-0225759.
XX   18-AUG-2000; 2000US-0226279.
XX   22-AUG-2000; 2000US-0226681.
XX   22-AUG-2000; 2000US-0226688.
XX   23-AUG-2000; 2000US-0227182.
XX   23-AUG-2000; 2000US-0227709.
XX   30-AUG-2000; 2000US-0228824.
XX   01-SEP-2000; 2000US-0229287.
XX   01-SEP-2000; 2000US-0229343.
XX   01-SEP-2000; 2000US-0229344.
XX   01-SEP-2000; 2000US-0229345.
XX   05-SEP-2000; 2000US-0229509.
XX   05-SEP-2000; 2000US-0229513.
XX   06-SEP-2000; 2000US-0230437.
XX   08-SEP-2000; 2000US-0231242.
XX   08-SEP-2000; 2000US-0231243.
XX   08-SEP-2000; 2000US-0231244.
XX   08-SEP-2000; 2000US-0231413.
```

```
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   25-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   01-NOV-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
```

```

PR   17-NOV-2000; 2000US-0249289.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251889.
PR   08-DEC-2000; 2000US-0251989.
PR   11-DEC-2000; 2000US-0251997.
PR   05-JAN-2001; 2001US-0259678.
PA   (HUMA-) HUMAN GENOME SCI INC.
PT   Rosen CA, Barash SC, Ruben SM.
DR   WPI; 2001-483426/52.
XX   Disclosure; SEQ ID NO 35725; 3071pp + Sequence Listing; English.
PS   AA
XX   AAKS4951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM821170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87654 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX   SO Sequence 37783 BP; 9418 A; 9546 C; 9545 G; 9274 T; 0 other:
XX   X
Alignment_scores:
      Quality:    81.00          Length:    39
      Ratio:      2.793         Gaps:      0
Percent Similarity: 74.359     Percent Identity: 46.154
alignment_block:
US-08-973-363714 x AAK80913/rev ..
Align seg 1/1 to reverse of: AAK80913 from: 1 to: 37783
       7  LysHSIPRTHSLySLySLleLySLaGLySLySLaSLngLutLSySLs 23
16327  AAGGAAAAAAGAAAGGAAGAAGAAAGGAAGAAAGGAAGGAAGGA 16278
       23  pGIuPrOGluLeGLyTleLySLySGluLaJGLyGluLSySLarGSLuThc 40
16277  GGAGGAAGAGGAGGAGGAGGAAGAGGAGGAGGAGGAGGAGGAGGA 16228
       40  ySGluLySLaSLnSLyS 45
       |||||||.....
16227  AGGAGAAAGGAAGAGGAG 16211
seq_name: /SIDSI/gcgdata/hold_geneseq/geneseqn_emb1/NA2001A.DAT.ABA446373

```

```

seq_documentation block:
ID      ABA46373 standard; DNA; 420 BP.
AC
XX      ABA46373:
DF      01-FEB-2002 (first entry)
DE      Human breast cell single exon nucleic acid probe #5008.
DX
XX      Human; microarray; single exon probe; gene expression; breast;
KM      disease; cancer; ss.
XX
OS      Homo sapiens.
XX
XX      WO200157271-A2.
PN
PD      09-AUG-2001.
PP
PF      30-JAN-2001; 2001WO-US00662.
PR
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
PA
PM      (MOLE-) MOLECULAR DYNAMICS INC.
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
DN      WPtI_2001-496933/54.
XX
XX      New spatially-addressable set of single exon nucleic acid probes,
PT      useful for measuring gene expression in sample derived from human
PT      breast. Comprises number of single exon nucleic acid probes -
PS
PS      Claim 1; SEQ ID NO 5068; 327bp + sequence listing; English.
XX
XX      The invention relates to a spatially-addressable set of single exon
CC      nucleic acid probes for measuring gene expression in a sample derived
CC      from human breast and BT 474 cells. The method involves contacting
CC      the probes with a collection of detectably labelled nucleic acids
CC      derived from mRNA of human breast, and then measuring the label
CC      bound to each probe of the microarray. The probes are useful for
CC      verifying the expression of regions of genomic DNA predicted to
CC      encode proteins. They are useful for gene discovery, and for
CC      determining predisposition and/or prognosing breast disease. Gene
CC      expression analysis is useful for assessing the toxicity of chemical
CC      agents on cells. The microarray of this invention presents a far greater
CC      diversity of probes for measuring gene expression, with far less bias
CC      than expressed sequence tag microarrays. The method is suitable for
CC      rapid production of functional information from genomic sequence. The
CC      present sequence is a single exon nucleic acid probe of the invention.
CC      Note: The sequence data for this patent did not form part of the
CC      printed specification, but was obtained in electronic format directly
CC      from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX      Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other:
SQ
Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other:

alignment_scores:
Quality:    79.00          Length:    45
Ratio:      2.469          Gaps:      2
Percent Similarity: 71.111   Percent Identity: 42.222

alignment_block:
DS-08-973-363-14 x ABA46373 ..

Align seg 1/1 to: ABA46373 from: 1 to: 420
4 ValSetValIySHSProHts.....lYsVSIleYSaJaagIuny 17

```

```

228 ATAACTTAGCTGATCCTCATTTCCAGAGAAAGAGAGATTGAGACA 277
17 sgluasnglululysaspgiuprogliileglyllyleystylsglualag 34
278 GGAGAGAGAGAGAGAGAGATGAGAGAA.....AACAGAGAGAGACA 318
34 lylulysargglulthrlsglulysgluasnllys 45
319 AGGAGAAAGAGAGAGAGAAATGAGAGAGAGAGAGAG 353

```

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA56939

seq\_documentation\_block:

ID ABA56939 standard; DNA; 420 BP.

XX ABA56939;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #5244.

XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human foetal liver.

XX Claim 1; SEQ ID NO 5244; 639bp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

XX measuring human gene expression in a sample derived from human foetal

XX liver. The single exon nucleic acid probes may be used for predicting,

XX measuring and displaying gene expression in samples derived from human

XX foetal liver. The present sequence is a single exon nucleic acid

XX probe of the invention.

XX Note: The sequence data for this patent did not form part of the

XX printed specification, but was obtained in electronic format directly

XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other;

alignment\_scores:

Quality: 79.00 Length: 45

Ratio: 2.469 Gaps: 2

Percent Similarity: 71.111 Percent Identity: 42.222

alignment\_block:

US-08-973-363-14 x ABA56939 ..

Align seg 1/1 to: ABA56939 from: 1 to: 420

```

4 ValserVallyshisprohis.....LysLysLleLysAlaGluLyl 17
228 ATAACTTAGCTGATCCTCATTTCCAGAGAGAGAGAGATTGAGACA 277
17 sgluasnglululysaspgiuprogliileglyllyleystylsglualag 34
278 GGAGAGAGAGAGAGAGATGAGAGAA.....AACAGAGAGAGACA 318
34 lylulysargglulthrlsglulysgluasnllys 45
319 AGGAGAAAGAGAGAGAGAAATGAGAGAGAGAGAGAG 353

```

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA26551

seq\_documentation\_block:

ID ABA26551 standard; DNA; 420 BP.

XX ABA26551;

XX 23-JAN-2002 (first entry)

XX Probe #5017 for gene expression analysis in human heart cell sample.

XX Human; gene expression; heart; microarray; vascular system; probe;

XX cardiovascular disease; hypertension; cardiac arrhythmia;

XX congenital heart disease; ss.

XX Homo sapiens.

XX WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX hearts.

XX Claim 1; SEQ ID NO 5017; 530bp; English.

XX The present invention relates to single exon nucleic acid probes for

XX measuring human gene expression in a sample derived from human heart. The

XX present sequence is one such probe. The probes may be used for

XX predicting, measuring and displaying gene expression in samples derived

XX from the human heart via microarrays. By measuring gene expression, the

XX probes are useful for predicting, diagnosing, grading, staging,

XX monitoring and prognosing diseases of the human heart and vascular system

XX e.g. cardiovascular disease, hypertension, cardiac arrhythmias and

XX congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other;

alignment\_scores:

Quality: 79.00 Length: 45

Ratio: 2.469 Gaps: 2

Percent Similarity: 71.111 Percent Identity: 42.222

alignment\_block:

US-08-973-363-14 x ABA26551 ..

Align seg 1/1 to: ABA26551 from: 1 to: 420

Ratio: 2.469 Gaps: 2  
Percent Similarity: 71.111 Percent Identity: 42.222  
Alignment\_block:  
US-08-973-363-14 x ABA26551 ..  
Align seg 1/1 to: ABA26551 from: 1 to: 420  
4 ValSerValLysHisProHis.....LysLysLleLysAlaGluLys 17  
:.....: ||||| ||||| ||| :|||  
228 ATAACTTACGTGATCCTCATTTCAAGGAGAGAGAGAGAGAGTTGAGAA 277  
17 sGUAenGluGluLysAspGluProGluLleGlyLleLysLysGluAlaG 34  
|||:.....: ||||| ||||| ||| :|||  
278 GGAGGAGAGAGAGAGAGATGATGAGAGAA.....AACAGAGAGAGA 318  
34 LysLysArGGLuThrLysGluLysGluAsnLys 45  
|||:.....: ||||| ||||| ||||| :|||  
319 AGGAGAGAGAGAGAGAAATGAGAGAGAGAGAGAG 353  
seq\_name: /STDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK05034  
seq\_documentation\_block:  
ID AAK05034 standard; DNA; 420 BP.  
XX AAK05034;  
XX AC  
XX DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe SEQ ID NO: 5025.  
XX KM Human; brain expressed exon; gene expression analysis; probe:  
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX OS epilepsy; cancer; ss.  
XX OS Homo sapiens.  
XX PN WC200157275-A2.  
XX PD  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001MO-US00667.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX DR WPI; 2001-483446/52.  
XX PT Single exon nucleic acid probes for analyzing gene expression in human  
XX PT brains -  
XX PS Example 4; SEQ ID NO: 5025; 650pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC brain. They can be used to measure gene expression in brain cell samples,  
XX CC which may enable the diagnosis and improved treatment of nervous system  
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX CC epilepsy and cancers. The present sequence is one of the probes of the  
XX CC invention.  
XX SQ Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other;

alignment\_scores:  
Quality: 79.00 Length: 45  
Ratio: 2.469 Gaps: 2  
Percent Similarity: 71.111 Percent Identity: 42.222  
Alignment\_block:  
US-08-973-363-14 x AAK05034 ..  
Align seg 1/1 to: AAK05034 from: 1 to: 420  
4 ValSerValLysHisProHis.....LysLysLleLysAlaGluLys 17  
:.....: ||||| ||||| ||| :|||  
228 ATAACTTACGTGATCCTCATTTCAAGGAGAGAGAGAGAGAGTTGAGAA 277  
17 sGUAenGluGluLysAspGluProGluLleGlyLleLysLysGluAlaG 34  
|||:.....: ||||| ||||| ||| :|||  
278 GGAGGAGAGAGAGAGATGATGAGAGAA.....AACAGAGAGAGA 318  
34 LysLysArGGLuThrLysGluLysGluAsnLys 45  
|||:.....: ||||| ||||| ||||| :|||  
319 AGGAGAGAGAGAGAGAAATGAGAGAGAGAGAGAG 353  
seq\_name: /STDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK30564  
seq\_documentation\_block:  
ID AAK30564 standard; DNA; 420 BP.  
XX AAK30564;  
XX AC  
XX DT 06-NOV-2001 (first entry)  
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 5121.  
XX KM Human; bone marrow expressed exon; gene expression analysis; probe:  
XX KW microarray; cancer; Leukemia; Lymphoma; myeloma; ss.  
XX OS Homo sapiens.  
XX PN WC200157276-A2.  
XX PD  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001MO-US00668.  
XX PR 04-FEB-2000; 2000US-0180312.  
XX PR 26-MAY-2000; 2000US-0207456.  
XX PR 30-JUN-2000; 2000US-0608408.  
XX PR 03-AUG-2000; 2000US-0632366.  
XX PR 21-SEP-2000; 2000US-0234687.  
XX PR 27-SEP-2000; 2000US-0236359.  
XX PR 04-OCT-2000; 2000GB-0024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX DR WPI; 2001-488900/53.  
XX PT Human genome-derived single exon nucleic acid probes useful for  
XX PT analyzing gene expression in human bone marrow -  
XX PS Example 4; SEQ ID NO: 5121; 658bp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC bone marrow. They can be used to measure gene expression in bone marrow  
XX CC samples, which may enable the improved diagnosis and treatment of cancers  
XX CC such as lymphoma, leukemia and myeloma. The present sequence is one of  
XX CC the probes of the invention.  
XX SQ Sequence 420 BP; 157 A; 42 C; 122 G; 99 T; 0 other;







xx The invention relates to a spatially-addressable set of single exon  
cc nucleic acid probes for measuring gene expression in a sample derived  
cc from human breast and BT 474 cells. The method involves contacting  
cc the probes with a collection of detectably labelled nucleic acids  
cc derived from mRNA of human breast, and then measuring the label  
cc bound to each probe of the microarray. The probes are useful for  
cc verifying the expression of regions of genomic DNA predicted to  
cc encode proteins. They are useful for gene discovery, and for  
cc determining predisposition and/or prognosing breast disease. Gene  
cc expression analysis is useful for assessing the toxicity of chemical  
cc agents on cells. The microarray of this invention presents a far greater  
cc diversity of probes for measuring gene expression, with far less bias  
cc than expressed sequence tag microarrays. The method is suitable for  
cc rapid production of functional information from genomic sequence. The  
cc present sequence is a single exon nucleic acid probe of the invention.  
cc Note: The sequence data for this patent did not form part of the  
cc printed specification, but was obtained in electronic format directly  
cc from WIPO at [ftp.wipo.int/pub/published\\_pvt\\_sequences](http://wipo.int/pub/published_pvt_sequences).  
xx

Sequence 452 BP; 168 A; 47 C; 130 G; 107 T; 0 other;

```
alignment_scores:
  quality: 79.00
  ratio: 2.469
  percent_similarity: 71.111
  length: 45
  gaps: 2
  percent_identity: 42.222
```

```
alignment_block:
US-08-973-363-14 x ABA43465 ..
```

Align seg 1/1 to: ABA43465 from: 1 to: 452

[illegible]

OM of: US-08-973-363-14 to: EST: \* out\_format: pfs  
 Date: Aug 3, 2002 3:03 AM

About: Results were produced by the Gencore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters: -DEV-x1h  
 -MODEL=framed.p2n.model  
 -O=/gen2\_1/USPTO\_spool/US08973363/runat\_01082002\_080049\_18570/bpp\_query.fasta\_1.638  
 -DB=EST -QPM=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.000  
 -MINMATCH=0.100 -LOOPEXT=0.000 -QGAPEXT=4.500  
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -YGAPEXT=6.000  
 -DELEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500 -DELOP=6.000  
 -LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
 -ALIGN=15 -MODE=LOCAL -OUTFMT=fts -NORM=ext -HEADSIZE=500  
 -MINLEN=0 -MAXLEN=200000000 -USER=US08973363\_@CSN1\_1.2938  
 -NCPU=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30  
 -NO\_XLPRY -WAIT -THREADS=1

Search information block:  
 Query: US-08-973-363-14  
 Query length: 45  
 Database: EST: \*  
 Database sequences: 13736207  
 Database length: -1841457050  
 Search time (sec): 4390.400000

Sequence	Strid	Orig	zScore	Escore	len	Documentation
gb_est1:AA748563	-	159.00	271.17	7.4e-06	343	AA748563 oa57f02.s1 NCI CGAP.GC
gb_est1:AA722163	-	159.00	269.68	8.9e-06	430	AT222163 qh81f02.s1 Soares.fsta
gb_est1:AA699918	-	159.00	269.30	9.4e-06	456	AA699918 s161f12.s1 Soares.fsta
gb_est1:AA6015597	-	159.00	266.70	3.1e-05	547	AA6015597 603641828f1 NIH.MGC.87
gb_est1:AA1890775	-	146.00	264.26	0.0002	821	AA1890775 w95f11.x1 NCI CGAP.Ut
gb_est1:BP239967	-	128.00	252.77	0.0037	246	BP239967 60105170f1 NIH.MGC.54
gb_est1:MA387264	-	90.00	154.54	0.0037	653	MA387264 MPT-S70088-101139-003
gb_gss:BM055497	-	85.50	149.12	4.6e-04	488	BM055497 RPT-24-278116.TJ RPT
gb_est1:BP068827	-	85.50	148.50	5.0e-04	537	BP068827 st03c11.y1 Gm-c1065.G1
gb_est1:BG780046	-	84.00	143.47	9.3e-04	796	BG780046 s8a9n05.y1 Gm-c1064.G
gb_est1:BE822458	-	83.00	143.38	51.11	294	BE822458 GM70001B10E2 Gm-T107G
gb_est1:BE211136	-	83.00	145.30	75.86	470	BE211136 s05n08.y1 Gm-c1019.G1
gb_est1:MA459832	-	83.00	144.80	80.86	507	MA459832 AL628267.XGC-gastrula
gb_gss:AC173980	-	82.00	139.37	162.28	905	AC173980 Pan troglodytes DNA, c
gb_gss:AC173980	-	81.50	147.35	58.33	237	AC173980 IM0284619R Mouse 10kb
gb_gss:BA270280	-	81.50	142.44	109.52	501	BA270280 CH230-13902.TV CHORZ-2
gb_gss:AA453048	-	81.00	139.76	154.30	665	AA453048 IM0251M2R Mouse 10kb
gb_gss:AA135720	-	81.00	138.01	193.26	869	AA135720 SP-0172_L22_F01.SPEE.St
gb_est1:BG288677	-	81.00	135.89	253.58	1200	BG288677 60238548f1 NIH.MGC.5
gb_est1:BF121943	-	80.50	141.54	122.84	448	BF121943 saf1d01.y3 Gm-c1019.G1
gb_est1:BM219469	-	80.50	141.23	127.90	470	BM219469 s188c02.y1 Gm-c1019.G1
gb_gss:BM11345	-	80.50	141.54	127.90	1196	BM11345 f17j12-77.1 IGF Arabid
gb_est1:BM219469	-	80.00	145.07	78.09	231	BM219469 C0925F01.3 NIG Mouse 1
gb_est1:AU162571	-	80.00	141.62	121.61	391	AU162571 AUL12571 Rice CDNA f1c
gb_gss:AA033021	-	80.00	139.32	163.30	555	AA033021 RPT-23-376P11.TV RPT
gb_est1:AT181274	-	80.00	138.70	176.82	610	AT181274 SWACAC32601SC Brugla
gb_est1:BG295549	-	80.00	137.39	209.22	745	BG295549 60239290f1 NIH.MGC.94
gb_est1:BF384761	-	80.00	135.50	253.30	935	BF384761 60204653f1 NCI CGAP.1
gb_est1:BM415214	-	80.00	135.64	261.94	973	BM415214 OP20286 Mixed Stage ES
gb_est1:AM201364	-	79.50	140.35	143.09	419	AM201364 sf02c04.y1 Gm-c1027.G1
gb_est1:BE805692	-	79.50	140.34	143.38	420	BE805692 s848e02.y1 Gm-c1061.G1
gb_gss:AA0834195	-	79.50	139.99	149.96	443	AA0834195 RS.5413.B2.G09.SPEE.RF
gb_est1:AM762170	-	79.50	139.45	160.71	481	AM762170 SM02d09.y1 Gm-c1027.G1
gb_est1:BG028923	-	79.50	136.14	245.57	796	BG028923 60229933f1 NIH.MGC.86
gb_est1:BG834011	-	79.00	139.58	157.89	416	BG834011 3520923 MARC LPHG Sus.8
gb_gss:AA076468	-	79.00	138.97	170.87	456	AA076468 RPT-23-376P10.TV RPT
gb_gss:AA076468	-	79.00	138.97	170.87	457	AA076468 RPT-23-376P10.TV RPT
gb_gss:AA186779	-	79.00	135.82	255.80	738	AA186779 SP_1007_B2.D12.T7A.St

seq_name	gb_est1:AA748563	seq_name	gb_est1:AA748563	seq_name	gb_est1:AA748563	seq_name	gb_est1:AA748563
gb_est1:BP974760	+	79.00	135.63	262.20	760	BP974760 60224540f1 NIH.MGC	
gb_gss:AG061263	+	79.00	135.59	263.36	764	AG061263 Pan troglodytes DNA	
gb_est1:BG847842	+	79.00	134.88	288.67	852	BG847842 1024019C11.x2.C.re	
gb_gss:CHS03275	+	79.00	133.07	364.20	1123	CHS03275 Tetrahodon nigrovitt	
gb_est1:BG026974	+	79.00	132.09	412.75	1303	BG026974 60229400f1 NIH.MG	

seq\_name: gb\_est1:AA748563

seq\_documentation\_block:

LOCUS AA748563 343 bp mRNA linear EST 27-JAN-1998

DEFINITION AA748563 Homo sapiens CDNA clone IMAGE:1309072.3

PROTEIN CHD-1. [1]: contains element 11 repetitive element ;, mRNA

ACCESSION AA748563

VERSION AA748563.1 GI:2788521

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 343)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.

CGAP Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

CGAP Library Arrayed by: Greg Lennon, Ph.D.

DNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. consortium/LLNL at: www-bio.llnl.gov/bhrp/image/html

Insert length: 1312. Str Error: 0.00

Seq primer: -10m13 fwd RT from Amersham

High quality sequence stop: 282.

Location/Qualifiers

1..343 /organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1309072"

/clone\_id="NCI.CGAP.GCBI"

/tissue\_type="germlinal center B cell"

/lab\_host="DH10B"

/note="Vector: pYT3D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand CDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI). Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). CDNA synthesis was primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCGAAGGCGGCGCCGCTCATTTTTTTTTTTTTTTT-3']

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pYT3D vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 62 a 69 c 35 g 177 t

ORIGIN

alignment\_scores: 159.00 Length: 45

Ratio: 4.077 Gaps: 0

Percent Similarity: 86.667 Percent Identity: 68.889

alignment\_block: US-08-973-363-14 x AA748563/rev ..





```
17 sgluAsnglululysaspgluprogliileglytlelysluag 34
      ::::::::::::::::::::
82 AGACAGTGAAGAAAAACCTGACGACATGTTATATAAGAAAGAACAG 33
34 glylulysargglutlulysglulysluasn 44
      ::::::::::::::::::::
32 AAGAAAAGAGGGAAGCAAAAAGAGAGAAAT 1

seq_name: gb_est2:BF239967

seq_documentation block:
LOCUS       BF239967               821 bp    mRNA           linear   EST 14-NOV-2000
DEFINITION  601905170F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4133129 5',
ACCESSION   BF239967
VERSION     BF239967.1  GI:11153890
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 821)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaabs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LINL at:
            http://image.llnl.gov
            Plate: L1CM1033 row: k column: 18
            High quality sequence stop: 562.
FEATURES
    source          location/Qualifiers
1..821
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="IMAGE:4133129"
    /clone_lib="NIH_MGC_54"
    /tissue_type="from chronic myelogenous leukemia"
    /lab_host="DH10B (T1 phage-resistant)"
    /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
    Site 1: Sfil (ggcgccgcgcgcgc); Site 2: Sfil (ggcgccatggcc
    ); Double-stranded cDNA was prepared from cell line RNA.
    5' and 3' adaptors were used in cloning as follows: 5'
    adaptor sequence: 5'-CACGGCCATTATGCGC-3' and 3' adaptor
    sequence: 5'-ATTCFAGAGCGCGACGCGACATG-dT(30)BN-3'
    (where B = A, C, G and N = A, C, G, or T). Average
    insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
    contained inserts by PCR. This library was enriched for
    full-length clones and was constructed by Clontech
    Laboratories (Palo Alto, CA)."
```

```
17 sgluAsnglululysaspgluprogliileglytlelysluag 33
      ::::::::::::::::::::
284 AGACAGTGAAGAAAAACCTGACGACATGTTATATAAGAAAGAACCA 333
34 glylulysargglutlulysglulysluasn 45
      ::::::::::::::::::::
334 GAAGAAAAGAGGGAAGCAAAAAGAGAGAAAT 369

seq_name: gb_est1:AW387264

seq_documentation block:
LOCUS       AW387264               249 bp    mRNA           linear   EST 04-FEB-2000
DEFINITION  MRL-ST0088-101199-003-b09 ST0088 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW387264
VERSION     AW387264.1  GI:6891923
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 249)
AUTHORS     HGP http://www.ludwig.org.br/ORESTES
TITLE       The FAPESP/LICR Human Cancer Genome Project
JOURNAL     Unpublished (1999)
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?l=MRL&t=MRL-ST0088-
            101199-003-b09&t3=1999-11-10&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 27
            High quality sequence stop: 248.
FEATURES
    source          location/Qualifiers
1..249
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone_lib="ST0088"
    /dev_stage="Adult"
    /note="Organ: stomach; Vector: puc18; Site 1: SmaI;
    Site 2: SmaI; A mini-library was made by cloning products
    derived from ORESTES PCR (U.S. Letters Patent application
    No. 196,716 - Ludwig Institute for Cancer Research)
    profiles into the puc 18 vector. Reverse transcription of
    tissue mRNA and cDNA amplification were performed under
    low stringency conditions."
```

109 AGACAGTGAAGAAAAACCTGACCATGTTTATATTAAGAAGACCAAG 60  
 34 LysGluLysArgGluThrLysGluLysGluLysAnlys 45  
 |||||.....  
 59 AAGCAAGAGGGAGGACACAGCAAGCAAGACTCA 25  
 seq\_name: gb\_gss: BH055497

seq\_documentation\_block:  
 LOCUS BH055497 653 bp DNA linear GSS 18-JUL-2001  
 DEFINITION RPI-24-278116.TJ RPI-24 Mus musculus genomic clone RPI-24-278116  
 , DNA sequence.  
 ACCESSION BH055497  
 VERSION BH055497.1 GI:14862403  
 KEYWORDS GSS.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 653)  
 Zhao, S., Niernan, M., Malek, J., Shatsman, S., Akiret, B., Levins, M.,  
 Tsengaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
 Russell, D., de Jong, P. and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPI-24  
 Unpublished (1999)  
 Other GSSs: RPI-24-278116.TJ  
 CONTACT: Shaping Zhao  
 Department of Pathologic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@tigr.org). Clones may be purchased from BACPAC  
 Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end  
 Page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 278 row: 1 column: 16  
 Seq Primer: SP6  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1..653  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_id="RPI-24-278116"  
 /clone\_id="RPI-24"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /note="Vector: pTABAC1; Site\_1: BamHI; Site\_2: BamHI;  
 RPI-24 Mouse BAC Library produced by Pieter de Jong. The  
 library was cloned in the pTABAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."  
 BASE COUNT 255 a 101 c 172 g 125 t  
 ORIGIN

alignment\_scores: Quality: 90.00 Length: 39  
 Ratio: 3.333 Gaps: 0  
 Percent Similarity: 69.231 Percent Identity: 51.282

Alignment\_block:  
 US-08-973-363-14 x BH055497 ..  
 Align seg 1/1 to: BH055497 from: 1 to: 653

7 LysHisProHisLysLysIleLysAlaGluLysGluAnGluGluLysAs 23  
 |||||.....  
 140 AAGAAAGAAAGAAAGAAAGAAAGAAAGATGATGAATGAAGAAAGAA 189

23 pgiuprogliuileglyllylelyslgsluaglygluLysArgGluThrL 40  
 |||||.....  
 190 AAGAAAGAAAGATAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTGA 239  
 40 ysgLysGluLysAnlys 45  
 |||||.....  
 240 AAGAAAAAGAAAAACAAA 256  
 seq\_name: gb\_est2: BF068827

seq\_documentation\_block:  
 LOCUS BF068827 488 bp mRNA linear EST 06-DEC-2001  
 DEFINITION s103c11.y1 Gm-cl065 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:  
 Gm-cl065-237 5' similar to TR:003982 003982 HYPOTHETICAL 19.8 KD  
 PROTEIN. mRNA sequence.  
 ACCESSION BF068827  
 VERSION BF068827.1 GI:10845778  
 KEYWORDS EST.  
 SOURCE soybean.  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 Glycine  
 1 (bases 1 to 488)  
 Shoemaker, R., Keim, P., Vodka, L., Erpelting, J., Corvelli, V., Khana  
 A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,  
 Wylie, T., Underwood, K., Stepcie, M., Theising, B., Allen, M., Bowers  
 J., Peterson, B., Swaller, T., Gidbons, M., Page, D., Harvey, N., Schurk  
 R., Ritzer, E., Kohn, S., Shin, T., Jackson, T., Cardenas, M., McCann  
 R., Waterston, R. and Wilson, R.  
 Public Soybean EST Project  
 Unpublished (1999)  
 CONTACT: Shoemaker R/Public Soybean EST Project  
 Public Soybean EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available through: Reagen, Invitrogen Corp. 2130  
 South Memorial Parkway Huntville, AL 35801 For further information  
 call: (800)-553-4363 or contact via email: [cust@reagen.com](mailto:cust@reagen.com)  
 Insert Length: 822 Std Error: 0.00  
 High quality sequence stop: 414.

FEATURES  
 source Location/Qualifiers  
 1..488  
 /organism="Glycine max"  
 /db\_xref="taxon:3847"  
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl065-237"  
 /clone\_id="Gm-cl065"  
 /tissue\_type="germinating shoots"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:  
 XhoI. The cDNA library was constructed from mRNA isolated  
 germinating shoots of the cultivar Williams. The seeds  
 were allowed to germinate for 24 hours prior to being  
 cold stressed for 2 days at 4C. Complementary DNA was  
 synthesized from mRNA using a primer consisting of a  
 poly(dT) sequence with a XhoI restriction site. EcoRI  
 adapters were ligated to the blunt-ended cDNA fragments  
 followed by XhoI digestion. The cDNA fragments were  
 directionally cloned into the EcoRI-XhoI restriction site  
 of the pBluescript vector. The ligated cDNA fragments were  
 transformed into DH10B host cells (GibcoBRL). This library  
 was constructed in the laboratory of Dr. Randy  
 Shoemaker."

BASE COUNT 180 a 73 c 115 g 120 t  
 ORIGIN

alignment\_scores: Quality: 85.50 Length: 45

Ratio: 2.672 Gaps: 2  
Percent Similarity: 71.111 Percent Identity: 46.667

Alignment\_block:  
US-08-973-363-14 x BF068827 ..

Align seg 1/1 to: BF068827 from: 1 to: 488

```

6 VallyshisProHis.....LysylsilelysalagluLy 17
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
38 GTGAAGGAGCCTGAGAGAGAGAGAGATCCCAAGAACTAGAGCAAGAA 87
17 sgluansgluLyLysaspgluProgluilegLylylelysgluAag 34
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
88 AGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 125
34 lygluLyArggluThrlysgluLyLysgluAaslys 45
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
126 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 160

```

seq\_name: gb\_est2:BG790466

seq\_documentation\_block:

LOCUS BG790466 537 bp mRNA linear EST 29-NOV-2001  
DEFINITION sae69h05.y1 Gm-cl064 Glycine max cDNA clone GENOME SYSTEMS CLONE  
ID: Gm-cl064-3801 5' similar to TR:003982 003982 HYPOTHETICAL 19.8  
KD PROTEIN ; mRNA sequence.

ACCESSION BG790466  
VERSION BG790466.1 GI:14126028

KEYWORDS EST.  
SOURCE soybean.

ORGANISM

Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

REFERENCE 1 (bases 1 to 537)

AUTHORS

Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna  
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,  
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers  
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schuck  
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., Mccann  
R., Waterston,R. and Wilson,R.  
Public Soybean EST Project  
Unpublished (1999)

TITLE JOURNAL

COMMENT Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.edu

This clone is available through: Resgen, Invitrogen Corp. 2130

South Memorial Parkway Huntville, AL 35801 For further information

call: (800)-533-4363 or contact via email: ccu@resgen.com

High quality sequence stop: 421.

FEATURES

source

Location/Qualifiers

1..537

/organism="Glycine max"

/cultivar="Williams"

/db\_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl064-3801"

/clone\_1lb="Gm-cl064"

/tissue\_type="seedling epicotyls"

/dev\_stage="2 week old"

/lab\_host="DH10B"

/note="Vector: pBluescript II SK+; Site\_1: EcoRI; Site\_2:

XhoI; The cDNA library was constructed from mRNA isolated

from the epicotyls of 2 week old seedling for the cultivar

Williams. The seedlings were germinated in a growth

chamber, excised above the soil level, and the plants

were placed in a 100 ppm solution of auxin for 24 hours

prior to harvesting. Complementary DNA was synthesized

BASE COUNT 193 a 93 c 126 g 125 t  
ORIGIN

alignment\_scores:

Quality: 85.50 Length: 45

Ratio: 2.672 Gaps: 2

Percent Similarity: 71.111 Percent Identity: 46.667

alignment\_block:

US-08-973-363-14 x BG790466 ..

Align seg 1/1 to: BG790466 from: 1 to: 537

```

6 VallyshisProHis.....LysylsilelysalagluLy 17
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
343 GTGAAGGAGCCTGAGAGAGAGAGATCCCAAGAACTAGAGCAAGAA 392
17 sgluansgluLyLysaspgluProgluilegLylylelysgluAag 34
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
393 AGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 430
34 lygluLyArggluThrlysgluLyLysgluAaslys 45
  |||:|||||:|||||:|||||:|||||:|||||:|||||:
431 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465

```

seq\_name: gb\_est2:BE822468

seq\_documentation\_block:

LOCUS BE822468 796 bp mRNA linear EST 24-MAY-2001  
DEFINITION Gm700017B10E2 Gm-r1070 Glycine max cDNA clone Gm-r1070-6747 3',  
mRNA sequence.

ACCESSION BE822468  
VERSION BE822468.1 GI:10254702

KEYWORDS EST.

SOURCE soybean.

ORGANISM

Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

REFERENCE 1 (bases 1 to 796)

AUTHORS

Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,  
Erpelting,J., Rapp,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.

A Functional Genomics Program for Soybean (NSF 9872565)

Unpublished (1999)

Other\_ESTs: AW277969 corresponding to Gm-cl019-3195 (5')

Contact: Vodkin, L.O., PI, A Functional Genomics Program for

Soybean (NSF 9872565)

Lewin, H. A., Director, Keck Center for Comparative and Functional

Genomics

University of Illinois

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA

Tel: (217) 244-6147

Fax: (217) 333-4582

Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World

Parkway Circle St. Louis, Missouri 63134. For further information

call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)

427-3324 or contact: clones@genomesystems.com or info@genome

systems.com web site: www.genomesystems.com

Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

Location/Qualifiers

1..796

FEATURES

source



/organism="Glycine max"  
/db\_xref="taxon:3847"  
/clone="Gm-r1070-6747"  
/clone\_1lb="Gm-r1070"  
/note="The library Gm-r1070 is a sequence-driven, retracted set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig which were retracted to form library Gm-r1070. The cDNA clones of the retracted Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Kretzschmar, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/research/projects/Soybean/Index.html. Retraction was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/BioTech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the "OTHER EST" field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under "OTHER EST."

BASE COUNT    208 a    177 c    107 g    274 t    30 others

ORIGIN

alignment\_scores:  
Quality:    84.00    Length:    39  
Ratio:    2.897    Gaps:    1  
Percent Similarity:    74.359    Percent Identity:    48.718

alignment\_block:  
US-08-973-363-14 x BE822468/rev ..

Align seg 1/1 to reverse of: BE822468 from: 1 to: 796

seq\_name: qb\_est1:BE211196

seq\_documentation\_block:  
LOCUS    BE211196    294 bp    mRNA    linear    EST 04-DEC-2001  
DEFINITION    sc58805.y1.Gm-cl039 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: KD  
             Gm-cl039-2194.5, similar to TR:003982 003982 HYPOHYPICAL 19.8 KD  
ACCESSION    BE211196  
VERSION    BE211196.1  
KEYWORDS    protein, mRNA sequence.  
SOURCE    EST.  
ORGANISM    soybean.  
glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.  
1 (bases 1 to 294)  
Shomaker,R., Keim,P., Vodkin,L., Erpelidng,J., Coryell,V., Rhanna

```

TITLE
JOURNAL
COMMENT
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 285 1800
Fax: 314 286 1810
Email: estevenson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Hunttsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
Insert Length: 1211 Std Error: 0.00
High quality sequence stop: 245.
FEATURES
source
1..294
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1039-2194"
/clone_1lb="Gm-c1039"
/tissue_type="whole seedling without cotyledons"
/lab_host="DH10B"
/note="Vector: pBluescriptII SK+; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from 2 week old seedlings with the cotyledons removed at
the time of harvest. The seedlings for the cultivar Ogden
were grown in a growth chamber using germination paper.
Complementary DNA was synthesized from mRNA using a primer
consisting of a poly(dT) sequence with a XhoI restriction
site. EcoRI adapters were ligated to the blunt-ended cDNA
fragments followed by XhoI digestion. The cDNA fragments
were directionally cloned into the EcoRI-XhoI restriction
site of the pBluescript vector. The ligated cDNA fragments
were transformed into DH10B host cells (Gibco BRL). This
library was constructed by Dr. Randy Shoemaker."
BASE COUNT
130 a 49 c 78 g 37 t
ORIGIN
Alignment_scores:
Quality: 83.00 Length: 39
Ratio: 2.677 Gaps: 1
Percent Similarity: 79.487 Percent Identity: 46.154
alignment_block:
US-08-973-363-14 x BE211196 ..
Align seg 1/1 to: BE211196 from: 1 to: 294
7 LYSNLSPROHSLSYLSYLSLELYSLLEAGLYLSYLSGNGSLNGLYLSYAS 23
+++++||| ||||: ::|||:|||||:|||||:|||||:|||||:
40 GAGGAGCCCAAGAAAGAGAGCGCTGAGCCCAAAAGAGAGAGGAAAAAG 89
23 pGluPrGluL1EcLY1LYLSYLSG1LAAGLYLSYLSYARG1YUPLT 40
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
90 AGAGCCT.....AGGAAAGAAAGAGAGAGAGAGAGAGAGCA 127
40 yscLyuLysGluAsnLys 45
|||||:|||||:|||||
128 AGAAGAGAGAGAGAGAGAG 144
seq_name: gp_est1:AM459832
seq_documentation_block:
LOCUS AM459832 470 bp mRNA linear EST 03-DEC-2001
DEFINITION sb95908.y1 Gm-c106 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-c1016-7959 5' similar to TR:003982 003982 HYPOTHETICAL 19.8 KD
PROTEIN:?, mRNA sequence.

```



## seq documentation block:

LOCUS AG173980 905 bp DNA linear GSS 09-JAN-2002  
 DEFINITION Pan troglodytes DNA, clone: RP43-044B16.TJ, genomic survey

ACCESSION AG173980  
 VERSION AG173980.1 GI:16703660  
 KEYWORDS GSS; GSS (genome survey sequence).  
 SOURCE Pan troglodytes male lymphocytes DNA, clone: RP43-044B16.TJ.  
 Male BAC Library clone: RP43-044B16.TJ.

## ORGANISM

Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

## REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

COMMENT

Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail: chimps@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,  
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library RPCI-43 This BAC  
 end was generated during the Rad process and may have higher chance  
 of clone tracking errors.

## PRIMERS

Sequencing: TJ

## LIBRARY

Vector : pBAC3.6

R.Site 1 : EcoRI

R.Site 2 : EcoRI

Location/Qualifiers

1..905

/organism="Pan troglodytes"

/db\_xref="taxon:9598"

/clone="RP43-044B16.TJ"

/sex="male"

/cell\_type="lymphocytes"

/clone\_id="RPCI-43 Chimpanzee Male BAC Library"

BASE COUNT 10 a 380 c 93 g 407 t 15 others

ORIGIN

## alignment\_scores:

Quality: 82.00 length: 35  
 Ratio: 3.154 gaps: 0  
 Percent Similarity: 74.286 Percent Identity: 54.286

## alignment block:

US-08-973-363-14 x AG173980/rev ..

Align seg 1/1 to reverse of: AG173980 from: 1 to: 905

```

11 LysLysIleLysGluLysGluLysAsnGluLysAspGluProGluIle 27
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
495 AAGCAAAAAAAAAAGAGAGAGAAACCCGAGAGGGGCGGAAAGGACA 446
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
27 eGlyIleLysGluLysGluLysArgGluThrLysGluLysGlu 44
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
445 AGGGGAAAAAAAAAGAGAGAGACACACACAAAGAGAAAAAAAA 396
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
44 sLys 45
   :|||
395 GGA 391

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 01:41:11 ; Search time 10310.5 Seconds  
(without alignments) 310.532 Million cell updates/sec

Title: US-08-973-363-2

Sequence: 1 ATCTTCAGATGATCTGTA.....CTCAGAGACTTGTGTGTCG 153

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb-ba: \*  
2: gb-hg: \*  
3: gb-in: \*  
4: gb-com: \*  
5: gb-ov: \*  
6: gb-pat: \*  
7: gb-ph: \*  
8: gb-pl: \*  
9: gb-pr: \*  
10: gb-ro: \*  
11: gb-sts: \*  
12: gb-sv: \*  
13: gb-un: \*  
14: gb-yi: \*  
15: em-ba: \*  
16: em-fun: \*  
17: em-hum: \*  
18: em-in: \*  
19: em-om: \*  
20: em-om: \*  
21: em-ov: \*  
22: em-pat: \*  
23: em-ph: \*  
24: em-pl: \*  
25: em-pl: \*  
26: em-ro: \*  
27: em-sts: \*  
28: em-un: \*  
29: em-yi: \*  
30: em-hg-hum: \*  
31: em-hg-inv: \*  
32: em-hg-other: \*  
33: em-hg-inv: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	153	100.0	153	6	A58683	A58683 Sequence 2
2	102	66.7	5349	10	MUSC83X	L10410 Mouse DNA-b
3	54	35.3	1311	6	AR029026	AR029026 Sequence
4	27	17.6	5947	9	AR006513	AR006513 Homo sapi
5	27	17.6	101220	9	AC092372	AC092372 Homo sapi
6	27	17.6	134365	9	AC012624	AC012624 Homo sapi
7	27	17.6	145025	2	AC021449	AC021449 Homo sapi
8	27	17.6	145639	2	AC008531	AC008531 Homo sapi
9	27	17.6	193446	2	AC091946	AC091946 Homo sapi
10	27	17.6	219258	2	AC022121	AC022121 Homo sapi
11	27	17.6	276181	2	AC092382	AC092382 Homo sapi
12	25	16.3	153	6	A58684	A58684 Sequence 3
13	25	16.3	2292	5	D14316	D14316 delta-cryst
14	25	16.3	6608	6	A58691	A58691 Sequence 10
15	25	16.3	6872	5	AF004397	AF004397 Gallus ga
16	21	13.7	136229	2	AC106847	AC106847 Homo sapi
17	20	13.1	153	6	A58685	A58685 Sequence 4
18	19	12.4	16797	4	MRHGGC	MRHGGC
19	19	12.4	81035	8	AB025631	AB025631 Arabidops
20	19	12.4	86606	2	AC097591	AC097591 Rattus no
21	19	12.4	98507	2	AC096218	AC096218 Rattus no
22	19	12.4	110000	2	LMFLCHR18_01	LMFLCHR18_01
23	19	12.4	112664	2	AC008400	AC008400 Homo sapi
24	19	12.4	154859	9	AC012596	AC012596 Homo sapi
25	19	12.4	156280	9	AC079595	AC079595 Homo sapi
26	19	12.4	156563	2	AC079458	AC079458 Homo sapi
27	19	12.4	160690	2	AC023457	AC023457 Homo sapi
28	19	12.4	160894	2	AL450462	AL450462 Homo sapi
29	19	12.4	167466	2	AC084117	AC084117 Homo sapi
30	19	12.4	175357	2	AC079775	AC079775 Homo sapi
31	19	12.4	182770	2	AL596210	AL596210 Homo sapi
32	19	12.4	190000	2	AC004580	AC004580 Homo sapi
33	19	12.4	198757	2	AC073661	AC073661 Homo sapi
34	19	12.4	202540	2	AC009483	AC009483 Homo sapi
35	19	12.4	206589	9	AC009483	AC009483 Homo sapi
36	19	12.4	210256	2	AC108116	AC108116 Homo sapi
37	19	12.4	210515	9	AC097382	AC097382 Homo sapi
38	19	12.4	210636	9	AC006443	AC006443 Homo sapi
39	19	12.4	215705	9	AL589931	AL589931 Human DNA
40	19	12.4	216438	2	AF060702	AF060702 Struthio
41	18	11.8	348	5	AF173269	AF173269 Scutigera
42	18	11.8	418	3	AF062893	AF062893 Scutigera
43	18	11.8	2266	9	BC017739	BC017739 Homo sapi
44	18	11.8	10591	1	AB006096	AB006096 Pasteurel
45	18	11.8	10591	1	AB006096	AB006096 Pasteurel

## ALIGNMENTS

RESULT	1	153 bp	DNA	Linear	PAT 06-MAR-1998
A58683	A58683	Sequence 2 from Patent WO9639505.			
LOCUS	A58683	153 bp	DNA	Linear	PAT 06-MAR-1998
DEFINITION	A58683	Sequence 2 from Patent WO9639505.			
ACCESSION	A58683	GI:3714246			
VERSION	A58683.1	GI:3714246			
KEYWORDS	unidentified.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 153)				
AUTHORS	Griffiths, R. and Tiwari, B.				
TITLE	AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS				
JOURNAL	Patent: WO 9639505-A-2 12-DEC-1996;				
COMMENT	ISIS INNOVATION (GB)				
FEATURES	Other publication AU 5906996 961224.				
source	Location/Qualifiers				
	1..153				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
BASE COUNT	61 a	37 c	27 g	28 t	
ORIGIN					



[illegible]

ORGANISM	Homo sapiens			
REFERENCE	Homo sapiens; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
ATTORNS	1 (bases 1 to 101220)			
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.			
REFERENCE	unpublished			
ATTORNS	2 (bases 1 to 101220)			
TITLE	DOE Joint Genome Institute.			
JOURNAL	Direct Submission			
REFERENCE	Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
ATTORNS	3 (bases 1 to 101220)			
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.			
JOURNAL	Direct Submission			
REFERENCE	Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
ATTORNS	On Dec 7, 2001 this sequence version replaced gi:15290448.			
TITLE	Draft Sequence Produced by DOE Joint Genome Institute			
COMMENT	www.jgi.doe.gov Finishing Completed at Stanford Human Genome Center www.shgc.stanford.edu Quality: Phrap Quality >=40 100% of Sequence; Estimated Total Number of Errors is 0. NOTE: This insert is not the entire sequence of the clone (entire sequence is 146,7kb). It is clipped at the overlap with AC012624. The number of bases overlapped is 90404.			
FEATURES	Location/Qualifiers			
SOURCE	1..101220			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="5"			
	/clone="RP11-58M12"			
BASE COUNT	34122 a 18662 c 17827 g 30409 t			
ORIGIN				
Query Match	17.6%;	Score 27;	DB 9;	Length 101220;
Best Local Similarity	100.0%;	Pred. No. 0.00013;		
Matches	27;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;
QY	76 CAGACCCGTCAGACTACCTCATTCAA 102			
Db	26887 CAGACCCGTCAGACTACCTCATTCAA 26861			
RESULT	6			
AC012624	134365 bp DNA linear PRI 21-JUL-2001			
LOCUS	Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.			
DEFINITION	AC012624			
ACCESSION	AC012624.6 GI:14993679			
VERSION	HTG.			
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 134365)			
ATTORNS	DOE Joint Genome Institute and Stanford Human Genome Center.			
TITLE	Direct Submission			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 134365)			
ATTORNS	DOE Joint Genome Institute.			
TITLE	Direct Submission			
JOURNAL	Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
REFERENCE	3 (bases 1 to 134365)			
ATTORNS	DOE Joint Genome Institute and Stanford Human Genome Center.			
TITLE	Direct Submission			
JOURNAL	Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
REFERENCE	4 (bases 1 to 134365)			
ATTORNS	DOE Joint Genome Institute and Stanford Human Genome Center.			

TITLE	Direct Submission
JOURNAL	Submitted (21-Jul-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES	Location/Qualifiers
SOURCE	1. 134365 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="5" /clone="CTD-2082117"
BASE COUNT	40414 a 24497 c 23503 g 43951 t
ORIGIN	
Query Match	17.6%; Score 27; DB 9; Length 134365;
Best Local Similarity	100.0%; Prod. No. 0.00013;
Matches	27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	76 CAGACCCGTGCAGACTACCTCATCAAA 102 
Db	118292 CAGACCCGTGCAGACTACCTCATCAAA 118318
RESULT	7
AC021449	143079 bp DNA linear HTG 10-SEP-2000
LOCUS	Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered pieces.
DEFINITION	AC021449
ACCESSION	AC021449.3 GI:10047806
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS	human.
SOURCE	Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 143079)
AUTHORS	Barren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL	Homo sapiens, clone RP11-58M12
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 143079)
JOURNAL	Barren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckler,R., Beda,F., Boguski,L., Bonkgalter,B., Brown,A., Burnett,G., Castle,A., Choepey,L., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dekrellano,K., Dewar,K., Domino,M., Doyle,M., Feneslor,U., Ferrelia,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K., McDonald,P., Marquis,N., McKean,P., Mccurt,A., McKernan,K., Mcneeters,R., Meldrim,J., Menus,L., Morrow,D., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Ollivar,T.M., Peterson,K., Plette,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and zody.M.
TITLE	Direct Submission
JOURNAL	Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 30 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Sep 10, 2000 this sequence version replaced gi:1407963.
FEATURES	All repeats were identified using RepeatMasker:
SOURCE	Smith,A.F.A. & Green, P. (1996-1997)
COMMENT	http://ftp.genome.washington.edu/RM/RepeatMasker.html
Center:	Whitehead Institute/ MIT Center for Genome Research
Center code:	WIBR
Web site:	http://www.seq.wi.mit.edu
Contact:	sequence.submissions@genome.wi.mit.edu
Project Information	
Center project name:	LS154
Center clone name:	58_M_12
Summary Statistics	

```

Sequencing vector: M13; M77815: 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 134743 bases at least 940
Consensus quality: 139227 bases at least Q30
Consensus quality: 140814 bases at least Q20
Insert size: 144000; agarose-fp
Insert size: 142179; sum-of-configs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-configs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record
* is arbitrary. Gaps between the contigs are represented
* runs of N, but the exact sizes of the gaps are unknown
* This record will be updated with the finished sequence
* as soon as it is available and the accession number
* be preserved.
*
* 1 38820: contig of 38820 bp in length
* 38821 38920: gap of 100 bp
* 38921 40411: contig of 1491 bp in length
* 40412 40511: gap of 100 bp
* 40512 43279: contig of 2768 bp in length
* 43280 43379: gap of 100 bp
* 43380 46905: contig of 3526 bp in length
* 46906 47005: gap of 100 bp
* 47006 51830: contig of 4825 bp in length
* 51831 51930: gap of 100 bp
* 51931 62619: contig of 10689 bp in length
* 62620 62719: gap of 100 bp
* 62720 75408: contig of 12689 bp in length
* 75409 75508: gap of 100 bp
* 75509 92516: contig of 17008 bp in length
* 92517 92616: gap of 100 bp
* 92617 106409: contig of 13793 bp in length
* 106410 106509: gap of 100 bp
* 106510 143079: contig of 36570 bp in length.
*
* Location/Qualifiers
*
*   source
*     1..143079
*       /organism="Homo sapiens"
*       /db_xref="taxon:9606"
*       /clone="RP11-58M12"
*       /clone_1fb="RP11 Human Male BAC"
*
*   misc-feature
*     1..38820
*       /note="assembly-fragment"
*       clone_end:SP6
*       vector_side:left"
*
*   misc-feature
*     38921..40411
*       /note="assembly-fragment"
*       40512..43279
*         /note="assembly-fragment"
*       43380..46905
*         /note="assembly-fragment"
*       47006..51830
*         /note="assembly-fragment"
*       51931..62619
*         /note="assembly-fragment"
*       62720..75408
*         /note="assembly-fragment"
*       75509..92516
*         /note="assembly-fragment"
*       92617..106409
*         /note="assembly-fragment"
*       106510..143079
*         /note="assembly-fragment"
*       clone_end:r7
*       vector_side:right"
*
* BASE COUNT  43971 a 26246 c 26678 g 45278 t 906 others
* ORIGIN

```



Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCCGTCAGACTACCTCATCAAA 102  
Db 116192 CAGACCCGTCAGACTACCTCATCAAA 116218

## RESULT 8

AC008531 145659 bp DNA linear HTG 14-FEB-2001  
LOCUS Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
DEFINITION 7 ordered pieces.

AC008531 3 GI:12830078  
VERSION HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
KEYWORDS human.  
SOURCE

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
2 (bases 1 to 145659)

REFERENCE 2 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

On Feb 14, 2001 this sequence version replaced gi:7528342.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

-----  
Project Information  
Center Project Name: 369535  
Center clone name: CIT-HSPC\_480B11

-----  
Summary Statistics  
Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 148000; pulse field gel estimation  
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 6.27 in Q20 bases; sum-of-contigs estimation

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

1 56174: contig of 56174 bp in length  
56175 56274: gap of unknown length  
56275 100874: contig of 44600 bp in length  
100875 100974: gap of unknown length  
100975 113127: contig of 12135 bp in length  
113128 113227: gap of unknown length  
113228 113327: gap of unknown length  
113328 113427: contig of 4963 bp in length  
113428 113527: gap of unknown length  
113528 113627: contig of 104 bp in length  
113628 113727: gap of unknown length  
113728 113827: contig of 3503 bp in length  
113829 123297: gap of unknown length  
123298 123398: contig of 22262 bp in length.

## FEATURES

1. 145659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

/clone="CIT-480B11"  
/clone\_1lb="Caltech human BAC library C"  
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others  
ORIGIN

Query Match 17.6%; Score 27; DB 2; Length 145659;  
Best Local Similarity 100.0%; Pred. No. 0.00013;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCCGTCAGACTACCTCATCAAA 102  
Db 46815 CAGACCCGTCAGACTACCTCATCAAA 46841

## RESULT 9

AC091946 193446 bp DNA linear HTG 09-JUN-2001  
LOCUS Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
DEFINITION PROGRESS \*\*\*, 33 unordered pieces.

AC091946  
VERSION AC091946.1 GI:14333882  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 193446)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
2 (bases 1 to 193446)

REFERENCE 2 (bases 1 to 193446)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

## COMMENT

-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

-----  
Project Information  
Center Project Name: 544799  
Center clone name: RP11-36012

-----  
Summary Statistics  
Consensus quality: 157767 bases at least Q40  
Consensus quality: 180259 bases at least Q30  
Consensus quality: 184175 bases at least Q20  
Estimated insert size: 204590; agarose-gel estimation  
Quality coverage: 5.38 in Q20 bases; sum-of-contigs estimation  
Quality coverage: 5.79 in Q20 bases; agarose-gel estimation  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 33 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1358: contig of 1358 bp in length  
1359 1458: gap of unknown length  
1459 2667: contig of 1209 bp in length  
2668 2669: gap of unknown length  
2670 2768: contig of 1056 bp in length  
2769 3823: gap of unknown length  
3824 3923: contig of 1997 bp in length  
3924 5920: gap of unknown length  
5921 6021: contig of 1474 bp in length  
6022 7494: gap of unknown length  
7495 9096: contig of 1502 bp in length

*	9097	9196:	gap of unknown length
*	9197	10260:	contig of 1064 bp in length
*	10261	10360:	gap of unknown length
*	10361	12460:	contig of 2100 bp in length
*	12461	12560:	gap of unknown length
*	12561	14611:	contig of 2051 bp in length
*	14611	14711:	gap of unknown length
*	14712	16381:	contig of 1670 bp in length
*	16382	16481:	gap of unknown length
*	16482	17968:	contig of 1487 bp in length
*	17969	18068:	gap of unknown length
*	18069	20434:	contig of 2366 bp in length
*	20435	20534:	gap of unknown length
*	20535	23515:	contig of 2981 bp in length
*	23516	23615:	gap of unknown length
*	23616	27563:	contig of 3948 bp in length
*	27564	27663:	gap of unknown length
*	27664	30987:	contig of 3324 bp in length
*	30988	31087:	gap of unknown length
*	31088	36065:	contig of 4978 bp in length
*	36066	36165:	gap of unknown length
*	36166	40978:	contig of 4813 bp in length
*	40979	41078:	gap of unknown length
*	41079	45663:	contig of 4585 bp in length
*	45664	45763:	gap of unknown length
*	45764	51745:	contig of 5982 bp in length
*	51746	51845:	gap of unknown length
*	51846	57359:	contig of 5514 bp in length
*	57360	57360:	gap of unknown length
*	57460	67881:	contig of 10422 bp in length
*	67882	67981:	gap of unknown length
*	67982	74132:	contig of 6151 bp in length
*	74133	74233:	gap of unknown length
*	74233	79795:	contig of 5563 bp in length
*	79796	79895:	gap of unknown length
*	79896	87511:	contig of 7616 bp in length
*	87512	87611:	gap of unknown length
*	87612	92791:	contig of 5180 bp in length
*	92792	92891:	gap of unknown length
*	92892	102794:	contig of 9903 bp in length
*	102795	102894:	gap of unknown length
*	102895	110866:	contig of 7972 bp in length
*	110867	110966:	gap of unknown length
*	110967	117571:	contig of 6605 bp in length
*	117572	123738:	gap of unknown length
*	123739	123838:	gap of unknown length
*	123839	130583:	contig of 6745 bp in length
*	130584	130683:	gap of unknown length
*	130684	141544:	contig of 10861 bp in length
*	141545	141644:	gap of unknown length
*	141645	169109:	contig of 27465 bp in length
*	169110	169209:	gap of unknown length
*	169210	193446:	contig of 24237 bp in length

```

FEATURES
source      location/Qualifiers
1. 193446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-36012"
/clone_lib="RPC1 human BAC library 11"
BASE COUNT  57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN

```

Query Match	17.6%	Score 27;	DB 2;	Length 193446;
Best Local Similarity	100.0%	Pred. No. 0.00013;		
Matches 27;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 76 CAGACCCGTCAGACTACCTCATCAA 102  
|||||  
Db 87703 CAGACCCGTCAGACTACCTCATCAA 87729

RESULT	10
AC022121/c	
LOCUS	219258 bp DNA linear PRI 30-AUG-2001
DEFINITION	Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
ACCESSION	AC022121
VERSION	AC022121.6 GI:15375145
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	1. (bases 1 to 219255)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2. (bases 1 to 219258)
AUTHORS	DOE Joint Genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	3. (bases 1 to 219258)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

4 (phases 1 to 219258)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission  
Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94596, USA

COMMENT  
On Aug 30, 2001 this sequence version replaced gi:1514108.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.4.

STS Content:  
WI-5811 G04974  
WI-13675 G23101  
SHGC-58345 G38487  
SHGC-103595 G57841.

FEATURES	Location/Qualifiers
source	1. .219258

```

ORGANISM="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CND-2007H13"
BASE COUNT 71954 a 42062 c 40933 g 64309 t
ORIGIN

```

	Query Match	17.6%	Score 27:	DB 9:	Length 219258;
	Best Local Similarity	100.0%;	Pred.	No. 0.00013;	
	Matches 27;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	76 CAGACCCGTCGAGACTACCTCATCAAA	102			
Db	89405 CAGACCCGTCGAGACTACCTCATCAAA	89379			

RESULT	11
AC092382/c	
LOCUS	AC092382
DEFINITION	Homo sapiens chromosome 5 clone RP11-75H1, 276181 bp DNA, linear HTG-03-JUL-2001, WORKING DRAFT SEQUENCE, 47 unordered pieces.

ACCESSION AC092382  
VERSION AC092382.1 GI:14589571  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 276181)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 276181)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>

Project Information  
Center Project Name: 435334  
Center Clone Name: RPO1-11\_75H1

Summary Statistics  
Consensus quality: 229677 bases at least Q40  
Consensus quality: 256163 bases at least Q30  
Consensus quality: 260799 bases at least Q20  
Estimated insert size: 271820; agarose-rip estimation  
Estimated insert size: 271820; sum-of-contigs estimation  
Quality coverage: 8.62 in Q20 bases; agarose-rip estimation  
Quality coverage: 5.35 in Q20 bases; sum-of-contigs estimation.  
NOTE: This is a working draft sequence. It currently consists of 47 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1061: contig of 1061 bp in length  
1062 1161: gap of unknown length  
1162 2827: contig of 1666 bp in length  
2828 2927: gap of unknown length  
2928 4227: contig of 1300 bp in length  
4228 4327: gap of unknown length  
4328 5618: contig of 1291 bp in length  
5619 5718: gap of unknown length  
5719 6983: contig of 1265 bp in length  
6984 7083: gap of unknown length  
7084 8423: contig of 1339 bp in length  
8423 8523: gap of unknown length  
8523 9771: contig of 1249 bp in length  
9772 9871: gap of unknown length  
9872 11198: contig of 1327 bp in length  
11199 11298: gap of unknown length  
11299 12414: contig of 1116 bp in length  
12415 12514: gap of unknown length  
12515 13806: contig of 1292 bp in length  
13807 13906: gap of unknown length  
13907 15360: contig of 1454 bp in length  
15361 15460: gap of unknown length  
15461 16667: contig of 1207 bp in length  
16668 16767: gap of unknown length  
16768 18390: contig of 1623 bp in length  
18391 18490: gap of unknown length  
18491 19541: contig of 1051 bp in length  
19542 21156: gap of unknown length  
21157 21256: contig of 1515 bp in length  
21257 22448: contig of 1192 bp in length  
22449 22548: gap of unknown length  
22549 24499: contig of 1951 bp in length  
24500 26384: gap of unknown length  
26385 26484: gap of unknown length  
26485 27883: contig of 1399 bp in length  
27884 27983: gap of unknown length

\* 27984 30386: contig of 2403 bp in length  
\* 30387 30486: gap of unknown length  
\* 30487 32480: contig of 1994 bp in length  
\* 32481 32580: gap of unknown length  
\* 32581 35441: contig of 2861 bp in length  
\* 35442 35541: gap of unknown length  
\* 35542 37800: contig of 2259 bp in length  
\* 37801 37900: gap of unknown length  
\* 37901 40587: contig of 2687 bp in length  
\* 40588 40687: gap of unknown length  
\* 40688 44159: contig of 3472 bp in length  
\* 44160 44260: gap of unknown length  
\* 44261 46336: contig of 2377 bp in length  
\* 46337 46736: gap of unknown length  
\* 46737 50082: contig of 3346 bp in length  
\* 50083 50182: gap of unknown length  
\* 50183 53988: contig of 3806 bp in length  
\* 53989 54088: gap of unknown length  
\* 54089 55592: contig of 2504 bp in length  
\* 55593 63352: gap of unknown length  
\* 63353 64254: contig of 4660 bp in length  
\* 64255 64353: gap of unknown length  
\* 64354 68825: contig of 4471 bp in length  
\* 68826 68925: gap of unknown length  
\* 68926 74393: contig of 5468 bp in length  
\* 74394 74493: gap of unknown length  
\* 74494 80561: contig of 6068 bp in length  
\* 80562 80661: gap of unknown length  
\* 80662 87626: contig of 6965 bp in length  
\* 87627 87726: gap of unknown length  
\* 87727 93600: contig of 5874 bp in length  
\* 93601 93701: gap of unknown length  
\* 93702 102024: contig of 8324 bp in length  
\* 102025 102125: gap of unknown length  
\* 102126 109905: contig of 7781 bp in length  
\* 109906 110005: gap of unknown length  
\* 110006 116737: contig of 6732 bp in length  
\* 116738 116837: gap of unknown length  
\* 116838 126897: contig of 9960 bp in length  
\* 126898 126997: gap of unknown length  
\* 126998 135114: contig of 8217 bp in length  
\* 135115 135215: gap of unknown length  
\* 135216 146713: contig of 11499 bp in length  
\* 146714 146813: gap of unknown length  
\* 146814 157894: contig of 11081 bp in length  
\* 157895 171752: gap of unknown length  
\* 171753 171852: contig of 13758 bp in length  
\* 171853 201687: gap of unknown length  
\* 201688 201787: contig of 29835 bp in length  
\* 201788 231269: contig of 29481 bp in length  
\* 231270 276181: gap of unknown length  
\* 276182 44813: contig of 44813 bp in length.

FEATURES  
source 1. 276181  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-75H1"

BASP COUNT 86859 a 51769 c 50026 g 82854 t 4673 others  
ORIGIN

Query Match 17.6% Score 27; DB 3; Length 276181;  
Best Local 27; Similarity 100.0%; Pident No. 0.00014;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 76 CAGACCGGTGCACACTACCTCATCAAA 102  
Db 213649 CAGACCGGTGCACACTACCTCATCAAA 213623

```

RESULT 12
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 3 from Patent WO9639505.
ACCESSION A58684
VERSION A58684.1 GI:3714247
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 153)
AUTHORS Griffiths,R. and Tiwari,B.
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;
COMMENT ISIS INNOVATION (GB)
FEATURES
source Location/Qualifiers
1..153
BASE COUNT 58 a 40 c 31 g 24 t
ORIGIN

Query Match 16.3%; Score 25; DB 6; Length 153;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCCGTCGAGACTACTCAT 98
Db 74 TACAGACCCGTCGAGACTACTCAT 98

RESULT 13
LOCUS D14316 2292 bp mRNA linear VRT 03-FEB-1999
DEFINITION delta-crystallin enhancer binding protein, complete sequence.
ACCESSION D14316
VERSION D14316.1 GI:391639
KEYWORDS ORF2.
SOURCE Gallus gallus (library: lambda gt11) 13 day embryo lens cDNA to mRNA, clone JF11.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianidae; Gallus.
REFERENCE 1 (bases 1 to 2292)
AUTHORS Funahashi,J.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel., Aging and
Cancer, Tohoku Univ., 4-1 Seiryō-machi, Aoba-ku, Sendai 980-77,
Japan (Tel:022-272-9499, Fax:022-272-3982)
2 (sites)
REFERENCE Funahashi,J., Sekido,R., Murali,K., Kamachi,Y. and Kondoh,H.
AUTHORS Delta-crystallin enhancer binding protein delta EFL is a zinc
TITLE finger-homodomain protein implicated in postgastrulation
embryogenesis
JOURNAL Development 119 (2), 433-446 (1993)
MEDLINE 9411644
REFERENCE 3 (bases 1 to 2292)
AUTHORS Funahashi,J.
JOURNAL Unpublished (1994)
FEATURES
source Location/Qualifiers
1..2292
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="lens"
/clone_lib="lambda gt11"
/dev_stage="13 day embryo"
257..1939
CDS

```

```

/note="beta-gal fusion protein binds to blocks 10 to 3 of
HN fragment of delta-crystallin enhancer."
/codon_start=1
/product="ORF2"
/protein_id="BA03262.1"
/db_xref="GI:391640"
/translaction="MELKKCNHCYLLKPPDNEFYKQDALQHLIRSSGKLLDLK
LIRLRGRNRVLIFSQVNMIDILAEIKYRQPPQRLDGIKGLRKAQDHFNAEG
SEDFCLSTFRAGGLGINLASADTVIPDSMDNPQDLQQAARHIGQKQVNIYRL
VTKGSVEEDILERAKKKMYLDHVIORMDTGTVYHTGSTSSSPRKKEELATLK
FGAEELFKPEBGEQEPQEMDIDELIKRAETRENEPGLTVGDELISQKVAFNSMD
EDDIELEPERNSNMWELLPEQRRRIIEERQKLELEITMLPRMNCARQJSEFNSSE
GRSRSRRYSGSDSDSITERRKRRKGRPTIIPRENIKGFSAELIRFKSYKKEGPF
LERLDVARDAEIYDSENDIRLGLVHNGCIKALKDSSGQERAGRLGKVGPTF
RISGVQVNAKIVISHEELAPLHKSIIPSPDEERKRVIVPCHTKAAFDIDMGKEDDSN
LVGIYEYAVGSMQCVNRSQSLNTEILLPDDPRIPRONSIRPVOTISLNT"
BASE COUNT 827 a 385 c 346 g 534 t
ORIGIN

Query Match 16.3%; Score 25; DB 5; Length 2292;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCCGTCGAGACTACTCAT 98
Db 1904 TACAGACCCGTCGAGACTACTCAT 1928

RESULT 14
LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 10 from Patent WO9639505.
ACCESSION A58691
VERSION A58691.1 GI:3714250
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 6608)
AUTHORS Griffiths,R. and Tiwari,B.
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL Patent: WO 9639505-A 10 12-DEC-1996;
COMMENT ISIS INNOVATION (GB)
FEATURES
source Location/Qualifiers
1..6608
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 2289 a 1207 c 1459 g 1653 t
ORIGIN

Query Match 16.3%; Score 25; DB 6; Length 6608;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCCGTCGAGACTACTCAT 98
Db 4123 TACAGACCCGTCGAGACTACTCAT 4147

RESULT 15
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the z chromosome
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete
cds.
ACCESSION AF004397
VERSION AF004397.1 GI:2501845
KEYWORDS
SOURCE chicken.

```



.....

---



```
PI Griffiths R, Tiwari B;
XX WPI: 1997-043127/04.
DR P-PSDB; AAM08146.
XX
PT Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determin. and to control sex of progeny
XX
PS Claim 8; Fig 3; 76pp; English.
XX
CC Bases 3655-3977 (AAT42756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (see also
CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene
CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).
CC Translated amino acid sequences of this region are provided in
CC AAM08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also
CC AAT42754-55) genes determine sex in birds and can be used to identify
CC the sex of an embryo, foetus etc. and to manipulate the sex of
CC progeny.
XX
SQ Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

Query Match          100.0%; Score 153; DB 18; Length 153;
Best Local Similarity 100.0%; Pred. No. 4.3e-67;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATTCCTTCAGATGATCCTGATATAAAACACAGCAAAACAGTTACGACCAAAAACCA 60
DB 1 attcttcagatgatctcgtataaaaaaccacacagcaaaacgttacgacacaaaaaccca 60
OY 61 CAAGCAAAACAGTTTACAGACCCCGTCGAGACTTACCTCTCAAACTACTTACGAGATCTT 120
DB 61 caagcaaaacagttacagaccgctgcagactactcactcaactactagcagatcctt 120
OY 121 GCAAAAAGAGAGCTCGAGACTTGTGTGCGC 153
DB 121 gcaaaaagagagctcagagacttgtgtgcg 153

RESULT 2
AAV59280
ID AAV59280 standard; cDNA; 1311 BP.
XX
AC AAV59280;
XX
DT 14-DEC-1998 (first entry)
XX
DE Altered telomere repeat binding factor 1 gene.
XX
KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;
KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT 1..1311
FT /*tag= a
FT /product= "A-TRF"
XX
XX WO9636066-A1.
XX
XX 20-AUG-1998.
XX
XX 13-FEB-1998; 98WO-US02765.
XX
XX 04-FEB-1998; 98US-0018628.
XX
XX 13-FEB-1997; 97US-0800264.
XX
XX (UVRQ ) UNITV ROCKEFELLER.
XX
XX Bianchi A, De Lange T, Van Steensel B;
PI
```

```
XX
XX WPI: 1998-480769/41.
DR P-PSDB; AAV59280.
XX
PT Nucleic acid encoding altered telomere repeat binding protein and
XX related vectors - transformants, hetero-dimers and antibodies, used
XX PT to inhibit shortening of telomerases caused by ageing or disease,
XX PT also used to extend life of cells in culture
XX
PS Claim 14; Page 110-111; 163pp; English.
XX
CC The altered vertebrate telomere repeat binding protein (A-TRF) has a
CC telomere repeat binding factor (TRF) dimerisation domain, and forms a
CC hetero-dimer with TRF, preventing it from binding to the specified repeat
CC sequence. A-TRF, optionally expressed by gene therapy, is used to
CC inhibit shortening of telomeres associated with ageing (for cosmetic
CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,
CC atrophy of the skin, age-related macular degeneration, atherosclerosis,
CC tumours and viral (including human immune deficiency virus) infection.
CC Cells expressing A-TRF also have an increased life span in vitro, e.g.
CC for expression of recombinant proteins or where intended for subsequent
CC transplant or for testing, eliminating the need for transformation.
XX
SQ Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

Query Match          35.3%; Score 54; DB 19; Length 1311;
Best Local Similarity 100.0%; Pred. No. 1.4e-17;
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 52 AAAAACCACAGCAAAACAGTTTACAGACCCGTCGAGACTTACTTCATCAACTA 105
DB 1258 aaaaaccacagcaaaacagttacagaccgctgcagactactcactcaactca 1311

RESULT 3
AAT42757
ID AAT42757 standard; DNA; 153 BP.
XX
AC AAT42757;
XX
DT 12-MAR-1997 (first entry)
XX
DE Chick CHD-1A gene fragment.
XX
KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KW CHD-1A; CHD-W; W chromosome; ss.
XX
OS Gallus sp.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
XX WO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96WO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX
XX P-PSDB; AAM08147.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX
```



PT birds - used for sex determ. and to control sex of progeny  
 XX  
 XX Claim 8; Fig 3; 76pp: English.  
 XX  
 CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757).  
 CC chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)  
 CC and the great tit CHD-W gene (see also AA142759). Translated amino  
 CC acid sequences of this region are provided in AA142756-49. The  
 CC CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes  
 CC determine sex in birds and can be used to identify the sex of an  
 CC embryo, foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;  
 Query Match 16.3%; Score 25; DB 18; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 0.0049;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 74 TACAGACCGTGCAGACTACCTCAT 98  
 ||||||||||||||||||  
 Db 74 TACAGACCGTGCAGACTACCTCAT 98  
 RESULT 4  
 AA142751  
 ID AA142751 standard; CDNA; 6608 BP.  
 XX  
 AC AA142751;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chicken CHD-1A gene.  
 XX  
 KM Bird; sex determination; chromodomain-helicase-DNA binding 1 Avian;  
 KM CHD-W; chromodomain-helicase-DNA binding on the W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 228..5390  
 FT /\*tag- a  
 XX  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PF 05-JUN-1996; 96MO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI; 1997-043127/04.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny  
 XX  
 PS Claim 1; Fig 5; 76pp: English.  
 XX  
 CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with  
 CC the closely related CHD-1A gene (AA142751) is suggested to initiate  
 CC female development in birds. The sequence of CHD-1A was deduced  
 CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
 CC library using a great tit CHD-W sequence (see also AA142755) as probe.  
 CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
 CC gene (see also AA142756-57). It is located on an autosome or Z  
 CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
 CC specific signal on hybridisation to genomic DNA of a non-fertile  
 CC bird and can be used for sex determin. of a bird. CHD-1A nucleic

CC acids can also be used to control the sex of the progeny of a bird.  
 XX  
 SQ Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;  
 Query Match 16.3%; Score 25; DB 18; Length 6608;  
 Best Local Similarity 100.0%; Pred. No. 0.0046;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 74 TACAGACCGTGCAGACTACCTCAT 98  
 ||||||||||||||||||  
 Db 4123 TACAGACCGTGCAGACTACCTCAT 4147  
 RESULT 5  
 AA13278/C  
 ID AA13278 standard; DNA; 13884 BP.  
 XX  
 AC AA13278;  
 XX  
 DT 19-MAR-1999 (first entry)  
 XX  
 DE Enterococcus faecalis genome contig SRQ ID NO:341.  
 XX  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
 KW vaccine; attenuation; computer readable medium; ds.  
 XX  
 OS Enterococcus faecalis.  
 XX  
 PN WO9650555-A2.  
 XX  
 PD 12-NOV-1998.  
 XX  
 PF 04-MAY-1998; 98MO-US08985.  
 XX  
 PR 14-NOV-1997; 97US-0066009.  
 PR 06-MAY-1997; 97US-0044031.  
 PR 16-MAY-1997; 97US-0046555.  
 XX  
 PA (HDMA-) HUMAN GENOME SCI INC.  
 XX  
 PT Barash SC, Dillon PJ, Kunsch CA;  
 XX  
 DR WPI; 1999-045171/04.  
 XX  
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 XX  
 PS Claim 1; Page 1453-1460; 2084pp; English.  
 XX  
 CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC AA132938 to AA13919 represent these nucleotide sequences which are  
 CC primary nucleotide sequences, also known as contigs. The computer-based  
 CC system can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 XX  
 SQ Sequence 13884 BP; 4353 A; 2723 C; 2370 G; 4422 T; 16 other;  
 Query Match 14.4%; Score 22; DB 20; Length 13884;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 TGATCCTGATATAAACACAA 33  
|||||  
DB 13595 TGATCCTGATATAAACACAA 13574

RESULT 6  
AAT42758  
ID AAT42758 standard; DNA; 153 BP.

XX AAT42758;

DT 12-MAR-1997 (first entry)

DE Chick CHD-W gene fragment.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
KM CHD-1A; CHD-W; W chromosome; ss.  
XX

OS Gallus sp.

XX Key Location/Qualifiers

FT misc\_difference 52..81

FT /tag- a

FT /note= "bases 52-81 are a repeat of bases 22-51  
and are ignored in the translated amino  
acid sequence given in Fig 3"

FT PN MO9639505-A1.

XX 12-DEC-1996.

PD 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX Griffiths R, Tiwari B;

XX WPI: 1997-043127/04.

DR P-PSDB; AAM08148.

XX Avian chromodomain-helicase-DNA binding genes determine sex in  
PT birds - used for sex determ. and to control sex of progeny

PS Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A - Avian) gene (AAT42757).  
CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
CC and the great tit CHD-W gene (AAT42759). Translated amino acid  
CC sequences of this region are provided in AAM08146-49. The CHD-1A  
CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
CC sex in birds and can be used to identify the sex of an embryo,  
CC foetus etc. and to manipulate the sex of progeny.

SO Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match 13.1%; Score 20; DB 18; Length 153;

Best Local Similarity 100.0%; Pred. No. 1.5;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 CAGTACAGACCCGTCAGA 89

DB 70 cagttacagaccctgcaga 89

RESULT 7  
AAS81107  
ID AAS81107 standard; cDNA; 803 BP.

AC AAS81107;

XX 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #16911.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001MO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.

DR P-PSDB; ABG16920.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID NO 16911; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [http://wipo.int/pub/published\\_pcl\\_sequences](http://wipo.int/pub/published_pcl_sequences).

SO Sequence 803 BP; 221 A; 195 C; 240 G; 147 T; 0 other;

Query Match 12.4%; Score 19; DB 23; Length 803;

Best Local Similarity 100.0%; Pred. No. 4.8;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 AACCAAGCAAAACAGTT 44

DB 173 aaccacaagcaaacagtt 191

RESULT 8  
AAF15189  
ID AAF15189 standard; cDNA; 905 BP.  
XX AAF15189;

XX 13-MAR-2001 (first entry)  
 DT Trichoderma reesei EST SEQ ID NO:7712.  
 DE  
 XX Multiple gene expression; filamentous fungal cell; EST;  
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
 KW culture condition; environmental stress; spore morphogenesis;  
 KW metabolic pathway engineering; catabolic pathway engineering; ss.  
 XX  
 OS Trichoderma reesei.  
 PN WO200056762-A2.  
 PD 28-SEP-2000.  
 XX 22-MAR-2000; 2000WO-US07781.  
 XX 22-MAR-1999; 99US-0273623.  
 XX (NOVO ) NOVO NORDISK BIOTECH INC.  
 PA (NOVO ) NOVO NORDISK AS.  
 XX  
 PI Berta RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
 DR MPI: 2000-594572/56.  
 PT Monitoring differential expression of genes in filamentous fungal cells  
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
 PT substrate of expressed sequence tags -  
 PS  
 PS Claim 89; Page 3111; 3161pp; English.  
 XX  
 CC The present invention describes a method for monitoring differential  
 CC expression of genes in a first filamentous fungal (FF) cell relative to  
 CC expression of the same genes in one or more second filamentous fungal  
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
 CC are used in the methods for monitoring differential expression of genes  
 CC in a first filamentous fungal (FF) cell relative to expression of the  
 CC same genes in one or more second filamentous fungal cells. Monitoring  
 CC the global expression of genes from FF cells allows the production  
 CC potential of the microorganisms to be improved. New genes may be  
 CC discovered, possible functions of unknown open reading frames can be  
 CC identified and gene copy number variation and stability can be  
 CC monitored. The expression of genes can be used to study how FF cells  
 CC adapt to changes in culture conditions, environmental stress, spore  
 CC morphogenesis, recombination, metabolic or catabolic pathway  
 CC engineering. Using ESTs provides several advantages over genomic or  
 CC random cDNA clones including elimination of redundancy as one spot on an  
 CC array equals one gene or open reading frame, and organisation of the  
 CC microarrays based on function of the gene products to facilitate  
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
 CC all specifically claimed in the present invention.  
 XX  
 XX Sequence 905 BP; 226 A; 265 C; 185 G; 218 T; 11 other:  
 Query Match 11.8%; Score 18; DB 21; Length 905;  
 Best local similarity 100.0%; Pred. No. 15;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 45 ACAGACCAAAAAACACA 62  
 DB 291 acagaccataaacacaca 308  
 RESULT 9  
 AAS81550

ID AAS81550 standard; cDNA; 5168 BP.  
 XX  
 AC AAS81550;  
 XX  
 DT 13-FEB-2002 (first entry)  
 DE  
 XX DNA encoding novel human diagnostic protein #17354.  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX  
 PI Dzmanac RT, Liu C, Tang YT;  
 DR MPI: 2001-639362/73.  
 DR P-PSDB; ABE17363.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS  
 PS Claim 1; SEQ ID No 17354; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 XX Sequence 5168 BP; 1428 A; 1397 C; 1194 G; 1149 T; 0 other:  
 Query Match 11.8%; Score 18; DB 23; Length 5168;  
 Best local similarity 100.0%; Pred. No. 15;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 5 TTCACATGATCCCGATA 22  
 DB 725 ttccagatgaccccgata 742  
 RESULT 10  
 AAK83089  
 ID AAK83089 standard; DNA; 225 BP.

XX AAK83089;  
AC  
XX 07-NOV-2001 (first entry)  
DT  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:37901.  
KW  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
PD  
XX 09-AUG-2001.  
PD  
XX  
PF 17-JAN-2001: 2001MO-US01354.  
XX  
XX 31-JAN-2000: 2000US-0179065.  
PR 04-FEB-2000: 2000US-0180628.  
PR 24-FEB-2000: 2000US-0184664.  
PR 02-MAR-2000: 2000US-0186350.  
PR 16-MAR-2000: 2000US-0189874.  
PR 17-MAR-2000: 2000US-0190076.  
PR 18-APR-2000: 2000US-0198123.  
PR 19-MAY-2000: 2000US-0205515.  
PR 07-JUN-2000: 2000US-0209467.  
PR 28-JUN-2000: 2000US-0214886.  
PR 30-JUN-2000: 2000US-0215135.  
PR 07-JUL-2000: 2000US-0216647.  
PR 07-JUL-2000: 2000US-0216880.  
PR 11-JUL-2000: 2000US-0217487.  
PR 11-JUL-2000: 2000US-0217496.  
PR 14-JUL-2000: 2000US-0218290.  
PR 26-JUL-2000: 2000US-0220963.  
PR 26-JUL-2000: 2000US-0220964.  
PR 14-AUG-2000: 2000US-0224518.  
PR 14-AUG-2000: 2000US-0224519.  
PR 14-AUG-2000: 2000US-0225213.  
PR 14-AUG-2000: 2000US-0225214.  
PR 14-AUG-2000: 2000US-0225266.  
PR 14-AUG-2000: 2000US-0225267.  
PR 14-AUG-2000: 2000US-0225268.  
PR 14-AUG-2000: 2000US-0225270.  
PR 14-AUG-2000: 2000US-0225447.  
PR 14-AUG-2000: 2000US-0225757.  
PR 14-AUG-2000: 2000US-0225758.  
PR 14-AUG-2000: 2000US-0225759.  
PR 18-AUG-2000: 2000US-0226279.  
PR 22-AUG-2000: 2000US-0226681.  
PR 22-AUG-2000: 2000US-0226688.  
PR 23-AUG-2000: 2000US-0227182.  
PR 23-AUG-2000: 2000US-0227009.  
PR 30-AUG-2000: 2000US-0228924.  
PR 01-SEP-2000: 2000US-0229287.  
PR 01-SEP-2000: 2000US-0229343.  
PR 01-SEP-2000: 2000US-0229344.  
PR 01-SEP-2000: 2000US-0229345.  
PR 05-SEP-2000: 2000US-0229509.  
PR 05-SEP-2000: 2000US-0229513.  
PR 06-SEP-2000: 2000US-0230437.  
PR 06-SEP-2000: 2000US-0230438.  
PR 06-SEP-2000: 2000US-0231242.  
PR 08-SEP-2000: 2000US-0231243.  
PR 08-SEP-2000: 2000US-0231244.  
PR 08-SEP-2000: 2000US-0231413.  
PR 08-SEP-2000: 2000US-0231414.  
PR 08-SEP-2000: 2000US-0232080.  
PR 08-SEP-2000: 2000US-0232081.  
PR 12-SEP-2000: 2000US-0231968.  
PR 14-SEP-2000: 2000US-0232397.  
PR 14-SEP-2000: 2000US-0232398.  
PR 14-SEP-2000: 2000US-0232399.  
PR  
PR 14-SEP-2000: 2000US-0232400.  
PR 14-SEP-2000: 2000US-0232401.  
PR 14-SEP-2000: 2000US-0233063.  
PR 14-SEP-2000: 2000US-0233064.  
PR 14-SEP-2000: 2000US-0233065.  
PR 21-SEP-2000: 2000US-0234223.  
PR 21-SEP-2000: 2000US-0234274.  
PR 25-SEP-2000: 2000US-0234997.  
PR 25-SEP-2000: 2000US-0234998.  
PR 25-SEP-2000: 2000US-0235484.  
PR 27-SEP-2000: 2000US-0235834.  
PR 27-SEP-2000: 2000US-0235836.  
PR 29-SEP-2000: 2000US-0236327.  
PR 29-SEP-2000: 2000US-0236367.  
PR 29-SEP-2000: 2000US-0236368.  
PR 29-SEP-2000: 2000US-0236369.  
PR 29-SEP-2000: 2000US-0236370.  
PR 02-OCT-2000: 2000US-0236802.  
PR 02-OCT-2000: 2000US-0237037.  
PR 02-OCT-2000: 2000US-0237038.  
PR 02-OCT-2000: 2000US-0237039.  
PR 02-OCT-2000: 2000US-0237040.  
PR 13-OCT-2000: 2000US-0239935.  
PR 13-OCT-2000: 2000US-0239937.  
PR 20-OCT-2000: 2000US-0240960.  
PR 20-OCT-2000: 2000US-0241221.  
PR 20-OCT-2000: 2000US-0241785.  
PR 20-OCT-2000: 2000US-0241786.  
PR 20-OCT-2000: 2000US-0241787.  
PR 20-OCT-2000: 2000US-0241808.  
PR 20-OCT-2000: 2000US-0241809.  
PR 20-OCT-2000: 2000US-0241826.  
PR 01-NOV-2000: 2000US-0244617.  
PR 08-NOV-2000: 2000US-0246474.  
PR 08-NOV-2000: 2000US-0246475.  
PR 08-NOV-2000: 2000US-0246476.  
PR 08-NOV-2000: 2000US-0246477.  
PR 08-NOV-2000: 2000US-0246478.  
PR 08-NOV-2000: 2000US-0246523.  
PR 08-NOV-2000: 2000US-0246524.  
PR 08-NOV-2000: 2000US-0246525.  
PR 08-NOV-2000: 2000US-0246526.  
PR 08-NOV-2000: 2000US-0246527.  
PR 08-NOV-2000: 2000US-0246528.  
PR 08-NOV-2000: 2000US-0246532.  
PR 08-NOV-2000: 2000US-0246609.  
PR 08-NOV-2000: 2000US-0246610.  
PR 08-NOV-2000: 2000US-0246611.  
PR 08-NOV-2000: 2000US-0246613.  
PR 17-NOV-2000: 2000US-0249207.  
PR 17-NOV-2000: 2000US-0249208.  
PR 17-NOV-2000: 2000US-0249209.  
PR 17-NOV-2000: 2000US-0249210.  
PR 17-NOV-2000: 2000US-0249211.  
PR 17-NOV-2000: 2000US-0249212.  
PR 17-NOV-2000: 2000US-0249213.  
PR 17-NOV-2000: 2000US-0249214.  
PR 17-NOV-2000: 2000US-0249215.  
PR 17-NOV-2000: 2000US-0249216.  
PR 17-NOV-2000: 2000US-0249217.  
PR 17-NOV-2000: 2000US-0249218.  
PR 17-NOV-2000: 2000US-0249244.  
PR 17-NOV-2000: 2000US-0249245.  
PR 17-NOV-2000: 2000US-0249264.  
PR 17-NOV-2000: 2000US-0249265.  
PR 17-NOV-2000: 2000US-0249297.  
PR 17-NOV-2000: 2000US-0249299.  
PR 17-NOV-2000: 2000US-0249300.  
PR 01-DEC-2000: 2000US-0250160.  
PR 01-DEC-2000: 2000US-0250391.  
PR 05-DEC-2000: 2000US-0251030.  
PR 05-DEC-2000: 2000US-0251988.  
PR 05-DEC-2000: 2000US-0256719.



PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241121.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-02539678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM.  
XX  
XX WPI; 2001-483426/52.  
XX  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX metastasis -  
XX

PS Disclosure; SEQ ID NO 37902; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 225 BP; 52 A; 55 C; 55 G; 63 T; 0 other;  
XX  
QY 125 AAGAGAGCTCAGAGA 141  
DB 19 aaagagagctcagaga 35  
IIIIIIIIIIIIIIIIIIII  
AAK90734  
ID AAK90734 standard; DNA; 225 BP.  
XX  
AC AAK90734;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen genomic sequence SEQ ID NO: 4310.  
XX  
XX Human: digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01324.  
XX  
PF 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.



```
Db      19 aaagagagctcagaga 35
|||||
RESULT 13
AAK90736
ID      AAK90736 standard; DNA: 225 BP.
XX
XX      AAK90736;
AC
XX
DT      05-NOV-2001 (first entry)
XX
XX      Human digestive system antigen genomic sequence SEQ ID NO: 4312.
DE
XX      Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW      ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW      digestive system disorder; Meckel's diverticulum; ds.
XX
XX      Homo sapiens.
OS
XX      WO200155314-A2.
PN
XX      02-AUG-2001.
PD
XX      17-JAN-2001; 2001WO-US01324.
PE
XX      31-JAN-2000; 2000US-0179065.
PR      04-FEB-2000; 2000US-0180628.
PR      24-FEB-2000; 2000US-0184664.
PR      02-MAR-2000; 2000US-0186350.
PR      16-MAR-2000; 2000US-0189874.
PR      17-MAR-2000; 2000US-0190076.
PR      18-APR-2000; 2000US-0198123.
PR      19-MAY-2000; 2000US-0205515.
PR      07-JUN-2000; 2000US-0209467.
PR      28-JUN-2000; 2000US-0214886.
PR      30-JUN-2000; 2000US-0215135.
PR      07-JUL-2000; 2000US-0216647.
PR      07-JUL-2000; 2000US-0216880.
PR      11-JUL-2000; 2000US-0217487.
PR      11-JUL-2000; 2000US-0217496.
PR      14-JUL-2000; 2000US-0218290.
PR      26-JUL-2000; 2000US-0220963.
PR      26-JUL-2000; 2000US-0220963.
PR      14-AUG-2000; 2000US-0224518.
PR      14-AUG-2000; 2000US-0224519.
PR      14-AUG-2000; 2000US-0225213.
PR      14-AUG-2000; 2000US-0225214.
PR      14-AUG-2000; 2000US-0225266.
PR      14-AUG-2000; 2000US-0225267.
PR      14-AUG-2000; 2000US-0225268.
PR      14-AUG-2000; 2000US-0225270.
PR      14-AUG-2000; 2000US-0225447.
PR      14-AUG-2000; 2000US-0225757.
PR      14-AUG-2000; 2000US-0225758.
PR      14-AUG-2000; 2000US-0225759.
PR      18-AUG-2000; 2000US-0226279.
PR      22-AUG-2000; 2000US-0226681.
PR      22-AUG-2000; 2000US-0226868.
PR      22-AUG-2000; 2000US-0227183.
PR      23-AUG-2000; 2000US-0227189.
PR      30-AUG-2000; 2000US-0228924.
PR      01-SEP-2000; 2000US-0229287.
PR      01-SEP-2000; 2000US-0229343.
PR      01-SEP-2000; 2000US-0229344.
PR      01-SEP-2000; 2000US-0229345.
PR      03-SEP-2000; 2000US-0229309.
PR      03-SEP-2000; 2000US-0229513.
PR      06-SEP-2000; 2000US-0230437.
PR      06-SEP-2000; 2000US-0230438.
PR      08-SEP-2000; 2000US-0231242.
PR      08-SEP-2000; 2000US-0231243.
PR      08-SEP-2000; 2000US-0231244.
PR      08-SEP-2000; 2000US-0231244.
```

```
PR      08-SEP-2000; 2000US-0231413.
PR      08-SEP-2000; 2000US-0231414.
PR      08-SEP-2000; 2000US-0232080.
PR      08-SEP-2000; 2000US-0232081.
PR      12-SEP-2000; 2000US-0231968.
PR      14-SEP-2000; 2000US-0232397.
PR      14-SEP-2000; 2000US-0233398.
PR      14-SEP-2000; 2000US-0233399.
PR      14-SEP-2000; 2000US-0233400.
PR      14-SEP-2000; 2000US-0233401.
PR      14-SEP-2000; 2000US-0233063.
PR      14-SEP-2000; 2000US-0233064.
PR      14-SEP-2000; 2000US-0233065.
PR      21-SEP-2000; 2000US-0234223.
PR      21-SEP-2000; 2000US-0234274.
PR      25-SEP-2000; 2000US-0234997.
PR      25-SEP-2000; 2000US-0234998.
PR      26-SEP-2000; 2000US-0235484.
PR      27-SEP-2000; 2000US-0235834.
PR      27-SEP-2000; 2000US-0235836.
PR      29-SEP-2000; 2000US-0236327.
PR      29-SEP-2000; 2000US-0236367.
PR      29-SEP-2000; 2000US-0236368.
PR      29-SEP-2000; 2000US-0236369.
PR      29-SEP-2000; 2000US-0236370.
PR      02-OCT-2000; 2000US-0236802.
PR      02-OCT-2000; 2000US-0237037.
PR      02-OCT-2000; 2000US-0237038.
PR      02-OCT-2000; 2000US-0237039.
PR      02-OCT-2000; 2000US-0237040.
PR      13-OCT-2000; 2000US-0239933.
PR      13-OCT-2000; 2000US-0239937.
PR      20-OCT-2000; 2000US-0241291.
PR      20-OCT-2000; 2000US-0241785.
PR      20-OCT-2000; 2000US-0241786.
PR      20-OCT-2000; 2000US-0241787.
PR      20-OCT-2000; 2000US-0241808.
PR      20-OCT-2000; 2000US-0241809.
PR      20-OCT-2000; 2000US-0241826.
PR      01-NOV-2000; 2000US-0244617.
PR      08-NOV-2000; 2000US-0244674.
PR      08-NOV-2000; 2000US-0246475.
PR      08-NOV-2000; 2000US-0246476.
PR      08-NOV-2000; 2000US-0246477.
PR      08-NOV-2000; 2000US-0246478.
PR      08-NOV-2000; 2000US-0246523.
PR      08-NOV-2000; 2000US-0246524.
PR      08-NOV-2000; 2000US-0246525.
PR      08-NOV-2000; 2000US-0246526.
PR      08-NOV-2000; 2000US-0246527.
PR      08-NOV-2000; 2000US-0246528.
PR      08-NOV-2000; 2000US-0246532.
PR      08-NOV-2000; 2000US-0246609.
PR      08-NOV-2000; 2000US-0246610.
PR      08-NOV-2000; 2000US-0246611.
PR      08-NOV-2000; 2000US-0246613.
PR      17-NOV-2000; 2000US-0249207.
PR      17-NOV-2000; 2000US-0249213.
PR      17-NOV-2000; 2000US-0249214.
PR      17-NOV-2000; 2000US-0249215.
PR      17-NOV-2000; 2000US-0249216.
PR      17-NOV-2000; 2000US-0249217.
PR      17-NOV-2000; 2000US-0249218.
PR      17-NOV-2000; 2000US-0249244.
PR      17-NOV-2000; 2000US-0249245.
PR      17-NOV-2000; 2000US-0249264.
PR      17-NOV-2000; 2000US-0249265.
```



PR 17-NOV-2000: 2000US-0249297.  
 PR 17-NOV-2000: 2000US-0249298.  
 PR 17-NOV-2000: 2000US-0249300.  
 PR 01-DEC-2000: 2000US-0250160.  
 PR 01-DEC-2000: 2000US-0250391.  
 PR 05-DEC-2000: 2000US-0251030.  
 PR 05-DEC-2000: 2000US-0251988.  
 PR 05-DEC-2000: 2000US-0256719.  
 PR 06-DEC-2000: 2000US-0251856.  
 PR 08-DEC-2000: 2000US-0251868.  
 PR 08-DEC-2000: 2000US-0251869.  
 PR 08-DEC-2000: 2000US-0251989.  
 PR 11-DEC-2000: 2000US-0251990.  
 PR 05-JAN-2001: 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI: 2001-502630/55.  
 XX  
 PT Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognosing disorders of the  
 PT digestive system, particularly cancer and cancer metastases -  
 PS  
 CC Disclosure: SEQ ID NO 4312; 986bp; English.  
 CC  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer, peckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a genomic DNA fragment  
 CC encoding a digestive system antigen of the invention.  
 CC  
 CC Sequence 225 BP: 52 A; 55 C; 55 G; 63 T; 0 other:  
 SQ  
 Query Match 11.1%; Score 17; DB 22; Length 225;  
 Best Local Similarity 100.0%; Pred. NO. 49;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 125 AAGAGAGGCTCAGAGA 141  
 ||||||||||||||||  
 DB 19 aaagagagctcagaga 35  
 RESULT 14  
 AAC05229  
 ID AAC05229 standard; cDNA: 311 BP.  
 XX  
 AC AAC05229;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein 5' EST, SEQ ID NO: 9304.  
 XX  
 KM Human: 5' EST: expressed sequence tag; secreted protein; cDNA isolation;  
 KM gene therapy; chromosome mapping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000: 2000EP-0200610.  
 XX  
 PR 26-FEB-1999: 9905-0122487.  
 XX  
 PA (GBST) GENSET.

XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX  
 DR WPI: 2000-500381/45.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 1; SEQ ID 9304; 71pp + CD-ROM; English.  
 XX  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.  
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be  
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used  
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors.  
 CC  
 CC Sequence 311 BP: 90 A; 74 C; 63 G; 84 T; 0 other:  
 SQ  
 Query Match 11.1%; Score 17; DB 21; Length 311;  
 Best Local Similarity 100.0%; Pred. NO. 48;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 130 GAGGCTCAGAGACTTG 146  
 ||||||||||||||||  
 DB 184 gagctcagagacttgg 200  
 RESULT 15  
 AAC79872/C  
 ID AAC79872 standard; cDNA: 657 BP.  
 XX  
 AC AAC79872;  
 XX  
 DT 09-FEB-2001 (first entry)  
 XX  
 DE Human secreted protein encoding cDNA for gene 24.  
 XX  
 KM Human: secreted protein; cytosolic; antiarthritic; antiasthmatic;  
 KM immunosuppressive; antiarteriosclerotic; antiinflammatory; nootropic;  
 KM neuroprotective; antidiabetic; tranquiliser; vulnereary; antibacterial;  
 KM antiproliferic; antiarhythmic; antirheumatic; cardiant; anti-HIV;  
 KM autoimmune disorder; allergic condition; cardiovascular disorder;  
 KM cancer; neurological disease; tissue repair; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200055176-A2.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 09-MAR-2000: 2000MO-US06057.  
 XX  
 PR 12-MAR-1999: 9905-0124142.  
 PR 11-JUN-1999: 9905-0138597.  
 PR 03-DEC-1999: 9905-0168666.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis G;  
 XX  
 DR WPI: 2000-638176/61.  
 DR P-PSDB; AABA4853.

XX Novel 49 human secreted proteins useful for diagnosis, prevention and  
 PT treatment of disorders including neurological, cell proliferative,  
 PT cardiovascular, and autoimmune/inflammatory disorders and microbial  
 PT infections -

XX Claim 1a: Page 349; 405pp: English.

XX  
 CC This invention describes a novel isolated polypeptide (I) comprising an  
 CC amino acid sequence at least 95 % identical to a polypeptide sequence  
 CC selected from 49 polypeptides encoded by polynucleotide sequences  
 CC included in American Type Culture Collection (ATCC) deposit number  
 CC 203917, defined in the specification. The products of the invention have  
 CC cytostatic, antiarthritic, antiasthmatic, immunosuppressive, nootropic,  
 CC antiarteriosclerotic, antiinflammatory, neuroprotective, antidiabetic,  
 CC tranquiliser, vulnery, antibacterial, antiporiatic, antiarrhythmic,  
 CC antineumatic, cardiant and anti-HIV activity. (I) or a nucleic acid (II)  
 CC encoding (I) is useful for preventing, treating or ameliorating a medical  
 CC condition and for diagnosing a pathological condition or susceptibility  
 CC to the condition. (I) is useful for identifying a binding partner which  
 CC affects the activity of the polypeptide and for identifying an activity  
 CC in a biological sample. (I), (II) or an antibody (IV) specific to (I) is  
 CC also useful for treating or preventing a disease, disorder or condition  
 CC associated with aberrant expression of (I). Diseases treated or diagnosed  
 CC include immune disorders such as autoimmune diseases, blood protein  
 CC disorders, anemia, allergic reactions and conditions such as asthma,  
 CC organ rejection or graft-versus-host disease, inflammation, hyper  
 CC proliferative disorders, cardiovascular disorders such as arterioarterial  
 CC fistula, arrhythmias, arteriosclerosis, coronary thrombosis, organ  
 CC regeneration, cancer, neovascular glaucoma, diabetic retinopathy,  
 CC rheumatoid arthritis, psoriasis, diseases associated with increased  
 CC apoptosis that include acquired immunodeficiency syndrome (AIDS),  
 CC neurological diseases such as Parkinson's disease, viral, bacterial,  
 CC fungal or parasitic diseases. They are also used to repair, replace or  
 CC protect tissue damage by congenital defects, to treat trauma, in surgery,  
 CC including cosmetic plastic surgery, to treat fibrosis, reperfusion injury,  
 CC or systemic cytokine damage, to stimulate chondrocyte growth, to prevent  
 CC skin aging due to sunburn, to change a mammal's mental state or physical  
 CC state by influencing biorhythms, cardiac rhythms, depression, memory,  
 CC stress and for accelerating wound healing. (I), (II) and/or their agonist  
 CC or antagonist are useful as food additives or preservatives to increase  
 CC or decrease storage capabilities, fat content, lipid, protein,  
 CC carbohydrate, vitamin, mineral or other nutritional components. (I) is  
 CC useful for screening therapeutic compounds. (II) is useful in forensic  
 CC biology for detecting DNA sequences and as diagnostic probes for  
 CC detecting the presence of specific mRNA in a particular cell type.  
 XX

SO Sequence 657 BP; 186 A; 125 C; 203 G; 143 T; 0 other;

Query Match 11.1%; Score 17; DB 21; Length 657;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 133 GCTCAGAGACTTTGTGG 149  
 ||||||||||||||||  
 Db 72 GCTCAGAGACTTTGTGG 56

Search completed: August 3, 2002, 01:50:02  
 Job time: 32187 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:29 ; Search time 7016.61 Seconds

(without alignments)  
294.307 Million cell updates/sec

Title: US-08-973-363-2

Perfect score: 153

Sequence: 1 ATGTCGACAGATGATCTGCA.....CTCAGAGACTTGTGCTGCG 153

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size: 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estlba:\*  
2: em\_estlhu:\*  
3: em\_estlin:\*  
4: em\_estlmu:\*  
5: em\_estlov:\*  
6: em\_estlpl:\*  
7: em\_estlro:\*  
8: em\_hic:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vtc:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	102	66.7	619	BB155356	BB155356 BB155356
2	102	66.7	660	BB461065	BB461065 BB461065
3	51	33.3	438	BB830730	BB830730 BB830730
4	35	22.9	446	BB834922	BB834922 BB834922
5	27	17.6	430	AL601246	AL601246 DKEB213J
6	27	17.6	547	AI890775	AI890775 wms5f11.x
7	27	17.6	686	AW997058	AW997058 OV3-BN004
8	27	17.6	821	BF239967	BF239967 601505170
9	27	17.6	866	AUI25712	AUI25712 AUI25712
10	27	17.6	1028	BB895133	BB895133 601436060
11	26	17.0	337	AW996787	AW996787 OV3-BN004
12	26	15.0	1200	AG131555	AG131555 Pan trogl
13	22	14.4	1312	BB858209	BB858209 1024056D0
14	22	12.4	217	BF147109	BF147109 UP07A12.Y
15	19	12.4	404	CNS02ED0	AL163107 Tetraodon
16	19	12.4	620	AI638930	AI638930 AEMTAD31
17	19	12.4	632	AZ848687	AZ848687 ZM0147B18

18	19	12.4	654	9	AI623066	AI623066 AEMTAD31
19	19	12.4	932	10	BB169355	BB169355 602320903
20	19	11.8	329	9	AA856136	AA856136 w83d08.r
21	18	11.8	380	9	AA148843	AA148843 z009e04.r
22	18	11.8	440	12	A0997521	A0997521 RPT-73-3
23	18	11.8	448	12	A0024759	A0024759 HS-2184.A
24	18	11.8	451	9	AI65697	AI65697 w83d08.Y
25	18	11.8	456	10	RA0931	RA0931 yf76a12.s1
26	18	11.8	481	9	AW634969	AW634969 b127c11.w
27	18	11.8	491	10	BU029333	BU029333 B0209393
28	18	11.8	495	12	BB570623	BB570623 B08X18TR
29	18	11.8	507	10	BC370667	BC370667 w73b06.Y
30	18	11.8	517	12	AO168710	AO168710 HS-3166.A
31	18	11.8	530	9	BB752282	BB752282 BB752282
32	18	11.8	545	10	BU068556	BU068556 BU068556
33	18	11.8	587	12	A098348	A098348 RPT-73-4
34	18	11.8	597	10	BU068189	BU068189 BU060189
35	18	11.8	635	10	BU043477	BU043477 BU043477
36	18	11.8	658	9	AL638257	AL638257 AL638257
37	18	11.8	658	9	AW639551	AW639551 B184C09.W
38	18	11.8	700	10	BG176934	BG176934 602313611
39	18	11.8	703	12	AG144450	AG144450 Pan trogl
40	18	11.8	709	10	A0545887	A0545887 CITR1-E1
41	18	11.8	810	10	BC666296	BC666296 DRABSA07
42	18	11.8	901	10	BE259673	BE259673 601145816
43	18	11.8	969	12	AG134136	AG134136 Pan trogl
44	18	11.8	993	12	CNS06G3K	AL397206 T3 end of
45	18	11.8	1176	10	BB036467	BB036467 602326413

## ALIGNMENTS

RESULT 1  
BB155356  
LOCUS  
DEFINITION  
BB155356 619 bp. mRNA linear. EST 18-OCT-2001  
musculus cDNA clone A13024116.3' similar to U10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION  
BB155356  
VERSION  
BB155356.2 GI:16268254

KEYWORDS  
KEYWORDS  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

REFERENCE  
1 (bases 1 to 619)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,M., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)

TITLE  
JOURNAL  
COMMENT  
On Jun 29, 2000 this sequence version replaced gi:8811286.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Sphero-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayashi,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagl,K., Fujikawa,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura



prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the C-terminal adenosine sequence CATATGCCCCCGCCC 3'. cDNA was cleaved with BamHI and XhoI. Vectors: a modified pluescript KS(+) after blunt excision from lambda FLIC I."

BASE COUNT					
ORIGIN	211	a	130	c	168 g 151 t

Query Match	66.7%;	Score 102;	DB 9;	Length 660;
Best Local Similarity	100.0%;	Pred. No. 3.1e-42;		
Matches 102; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	52	AAAAAACCAAGCAAGCAAAACAGTTACAGACCCGTGCAGACATACCTCAACAACTACTATGC	111
Db	502	AAAAAACCAAGCAAGCAAAACAGTTACAGACCCGTGCAGACATACCTCAACAACTACTATGC	561
QY	112	AGAGATCTTCGAAAAAGAGAGCGTCGAGACATTGTGTGTGCG	153
Db	562	AGAGATCTTCGAAAAAGAGAGCGTCGAGACATTGTGTGTGCG	603

RESULT	3	
LOCUS	BB830730	
DEFINITION	BB830730 438 bp mRNA linear EST 19-NOV-2001 BB830730 RIKEN full-length enriched, mammary gland cDNA JY9-MC(B) CDNA Mus musculus CDNA clone G930013K04 3', mRNA sequence.	
ACCESSION	BB830730	
VERSION	BB830730.1	GI:17008973
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE  
AUTHORS

Akimura, T., Arikawa, T., Carninci, P., Furuno, M., Hanegaki, T., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Okamura, M., Sakai, C., Sakai, K., Sakazume, N., Saeki, D., Sato, K., Saito, R., Shingagawa, A., Shikata, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takehashi, F., Takaku, Kikihara, S., Tanaka, T., Tomaru, A., Toya, T., Watanuki, A., Yasunishi, A., Yamatatsu, S. and Hayashizaki, I.

RRIEN Encyclopedia of Mouse Full-Length cDNAs (Akimura, T., et al. 2001)

JOURNAL  
COMMENT

Unpublished (2001)  
Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Sushiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp,  
URL: <http://genome-gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. *Genome Res.* 10 (10): 1617-1630 (2000)  
Wagti, K., Fujisake, S., Inoue, K., Nogawa, Y., Izawa, M., Ohara, F.,  
Matsubiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
, S., Kasai, T., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. *Genome Res.*  
10 (11): 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,  
Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA

FEATURES  
source  
1. .436  
Location/Qualifiers  
encyclopedia. real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2): 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.  
e mouse tissues.  
Location/Qualifiers  
1. .436

```

./db_xref="taxon:10090"
/clone="G930013K04"
/clone_1b="RIKEN full-length enriched, mammary gland
RCB-0527 Jyg-MC(B) CDNA"
/tissue_type="mammary gland"
/cell_line="RCB-0527 Jyg-MC(B)"
BASE COUNT      138 a      96 c      108 g      96 t
BRIGIN

```

Query Match	33.3%;	Score 51;	DB 9;	Length 438;
Best Local Similarity	100.0%;	Pred. No. 7.1e-16;		
Matches	51;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

Qy 1 ATTCCTCCAGATGATCCTGATATAAAACACACAGCAAAACAGTTACGACC 51  
|||||  
Dd 300 ATTCCTCCAGATGATCCTGATATAAAACACACAGCAAAACAGTTACGACC 350

RESULT	4				
LOCUS	BB834922				
DEFINITION	BB834922	446 bp	mRNA	linear	EST 19-NOV-2001
	BB834922				
	JY9-NC(B)				
	sequence.				
ACCESSION	BB834922.1	GI:17013165			
VERSION	EST.				
KEYWORDS	house mouse.				
SOURCE	Mus musculus				
ORGANISM					

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 446)				
2	Akimura, T., Arikawa, T., Carninci, P., Furuno, M., Hanagaki, T., Hayashizaki, Y., Hiramoto, K., Hiroaki, T., Hirozane, T., Imotani, K., Ishihara, T., Ito, M., Kawaji, Y., Kojima, Y., Kono, H., Kouda, M., Matsuyama, T., Nakamura, M., Nishikawa, K., Nomura, K., Nishizaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shirai, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akhiba, S., Tanaka, T., Tomaru, A., Toyo, T., Watanabe, K., Yasunishi, A., Yamamoto, M., and Hayashizaki, Y.	Genome-wide analysis of mouse full-length cDNAs	AKIMURA, T., et al. (2001)	Unpublished (2001) Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Sushiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp/  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1611-1630 (2000)  
Wagti, R., Fujiwara, S., Inoue, K., Togawa, Y., Izaara, M., Ohara, E., Matsubara, Y., Yoneda, Y., Ishikawa, K., Tanaka, T., Matsunaga, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
e mouse tissues.

FEATURES  
source  
Location/Qualifiers  
1..446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930033j21"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jyg-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jyg-MC(B)"

BASE COUNT  
139 a 99 c 108 g 100 t

Query Match 22.9%; Score 35; DB 9; Length 446;  
Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATCTTCAGATGATCCTGATTAACACACAGC 35  
|||||  
Db 309 ATCTTCAGATGATCCTGATTAACACACAGC 343

RESULT 5  
LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001  
DEFINITION DKFZ3131040.J1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 430)  
Bloecker, H., Boecker, M., Brandt, P., Mewes, W., Weill, B. and Wiemann, S.  
EST (Bloecker, H., Boecker, M., Brandt, P., Mewes, W., Weill, B. and Wiemann, S.)  
Unpublished (1999)  
Contact: Bloecker H  
MIPS  
Am Kioferplatz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
Braunschweig/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
No sl sequence available.  
This clone (DKFZ3131040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1..430

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZ3131040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pT7blue2; Site\_1: SfiI; Site\_2: SfiI; B;  
cDNA-collection  
163 a 81 c 86 g 100 t

BASE COUNT  
163 a 81 c 86 g 100 t

Query Match 17.6%; Score 27; DB 9; Length 430;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 CAGACCCGTCAGACTACCTCATCAAA 102  
|||||  
Db 263 CAGACCCGTCAGACTACCTCATCAAA 289

RESULT 6  
LOCUS AI890775 547 bp mRNA linear EST 07-MAR-2000  
DEFINITION wmg5f11.x1 NCI CGAP UT2 Homo sapiens cDNA clone IMAGE:2443725 3'  
similar to SW:CHD1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1; mRNA sequence.  
AI890775  
VERSION AI890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 547)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
[www.bio.lnl.gov/bbrp/image/image.html](http://www.bio.lnl.gov/bbrp/image/image.html)  
Insert length: 1924 Std Error: 0.00  
Seq primer: -40up from GIBCO  
High quality sequence stop: 418.  
Location/Qualifiers  
1..547

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2443725"  
/clone\_lib="NCI-CGAP-UT2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

BASE COUNT  
114 a 118 c 85 g 230 t

Query Match 17.6%; Score 27; DB 9; Length 547;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 CAGACCCGTCAGACTACCTCATCAAA 102  
|||||  
Db 333 CAGACCCGTCAGACTACCTCATCAAA 307

RESULT 7  
LOCUS AW997058 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION OV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW997058

VERSION AM97058.1 GI:8257292  
 EST.  
 KEYWORDS human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 686)  
 AUTHORS Nagai,M.A., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Dias Neto,E., Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=6t2-QV3-BN0047-150  
 400-152-c03&t3=2000-04-15&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 19  
 Location/Qualifiers  
 1..686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BN0047"  
 /dev\_stage="Adult"  
 /note="Organ: Breast;normal; Vector: puc18; Site:1: SmaI;  
 Site:2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 165 a 154 c 126 g 241 t  
 ORIGIN  
 Query Match 17.6%; Score 27; DB 9; Length 686;  
 Best Local Similarity 100.0%; Pred. No. 0.0018;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 76 CAGACCCGTCGACGACTACCTCATCAAA 102  
 ||||||||||||||||||||||||||||  
 Db 83 CAGACCCGTCGACGACTACCTCATCAAA 57  
 RESULT 8  
 LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000  
 DEFINITION 601905110F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
 mRNA sequence.  
 ACCESSION BF239967  
 VERSION BF239967.1 GI:11153890  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 821)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs@femail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 plate: LCM1033 row: k column: 18  
 High quality sequence stop: 562.  
 Location/Qualifiers  
 1..821  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:4133129"  
 /clone\_lib="NIH\_MGC\_54"  
 /issue\_type="from chronic myelogenous leukemia"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: bone marrow; Vector: pONR-LIB (Clontech);  
 Site:1: SfiI (ggccgctcgcc); Site:2: SfiI (ggccatctggcc  
 ); Double-stranded cDNA was prepared from cell line RNA.  
 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-ATCTAGAGCCGAGCGCCGCGCATG-3' and 3' adaptor  
 sequence: 5'-CAGGCGCATATGCGC-3' (where B = A, C, G, or T). Average  
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."  
 BASE COUNT 345 a 149 c 190 g 137 t  
 ORIGIN  
 Query Match 17.6%; Score 27; DB 10; Length 821;  
 Best Local Similarity 100.0%; Pred. No. 0.0018;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 76 CAGACCCGTCGACGACTACCTCATCAAA 102  
 ||||||||||||||||||||||||||||  
 Db 33 CAGACCCGTCGACGACTACCTCATCAAA 59  
 RESULT 9  
 LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
 DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
 sequence.  
 ACCESSION AU125712  
 VERSION AU125712.1 GI:10950428  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 866)  
 AUTHORS Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
 Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
 Isogai,T.  
 TITLE HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
 Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
 Y., Sugano,S., Isogai,T.)  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1537-3 Yana, Kisarazu, Chiba 297-0812, Japan  
 Tel: 81-438-52-3951  
 Fax: 81-438-52-3952  
 Email: genomics@hri.co.jp  
 HRI human cDNA project: 5'- & 3'-end one pass sequencing. Helix  
 Research Institute, cDNA library construction. Department of

Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.

FEATURES  
source

1. 866  
/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_11b="NT2RM4002061"

/clone\_11b="NT2RM4"

/cell\_type="teratocarcinoma"

/note="Vector: PME18SFL3; mRNA from uninduced NT2 neuronal

precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t 2 others

ORIGIN

Query Match 17.6%; Score 27; DB 9; Length 866;

Best Local Similarity 100.0%; Pred. No. 0.0018;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCCGTGCAGACTACCTCATCAAA 102

DB 495 CAGACCCGTGCAGACTACCTCATCAAA 521

RESULT 10

BE895133 1028 bp mRNA linear EST 20-OCT-2000

LOCUS 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',

DEFINITION mRNA sequence.

ACCESSION BE895133.1 GI:10358221

VERSION

KEYWORDS

SOURCE

ORGANISM

human.  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 1028)

NIH-MGC http://mhc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC/DCCTD/DTF

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLM9753 row: h column: 16

High quality sequence stop: 488.

Location/Qualifiers

1. 1028

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_11b="NT2RM4002061"

/clone\_11b="NT2RM4"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: Skin; Vector: PCMV-SPORE6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 2 kb. Library constructed by Life

Technologies."

Db 176 CAGACCCGTGCAGACTACCTCATCAAA 202

RESULT 11

AW996787

LOCUS 337 bp mRNA linear EST 05-JUN-2000

DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW996787

VERSION AW996787.1 GI:8257021

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 337)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordi, S., Costa, F.F.,

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and

Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-230

200-102-d03&tl=2000-02-23&tl=1)

Seq primer: puc 18 forward

High quality sequence start: 2

High quality sequence stop: 337.

Location/Qualifiers

1. 337

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_11b="BN0047"

/dev\_stage="Adult"

/note="Organ: breast, normal; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

ORIGIN

Query Match 17.0%; Score 26; DB 9; Length 337;

Best Local Similarity 100.0%; Pred. No. 0.0059;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 GATTAACACACACACCAACAGT 44

DB 87 GATTAACACACACACCAACAGT 62

RESULT 12

AG131555 1200 bp DNA linear GSS 04-NOV-2001

LOCUS AG131555

DEFINITION Pan troglodytes DNA, clone: PTB-143K14.F, genomic survey sequence.

ACCESSION AG131555

VERSION AG131555.1 GI:1661233



KEYWORDS GSS: GSS (genome survey sequence).  
SOURCE Pan troglodytes male lymphoblast DNA, clone\_11b:PTB Chimpanzee Male BAC Library clone:PTB-143K14.F.  
ORGANISM Pan troglodytes  
REFERENCE 1 (sites)  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokioka, Y., Watanabe, H. and Sakaki, Y.  
TITLE BAC end sequences of library PTB  
JOURNAL Unpublished  
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Tokioka, Y., Watanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC): 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpes@gsc.riken.go.jp, URL:http://ngp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170) ngs:chimpes:go.jp/  
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the Rsd process and may have higher chance of clone tracking errors.  
PRIMERS  
Sequencing: -21M13  
LIBRARY  
Vector : pRS145  
R.Site 1 : SacI  
R.Site 2 : SacI.  
Location/Qualifiers  
1. 1200  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone="PTB-143K14.F"  
/sex="male"  
/cell\_type="lymphoblast"  
/clone\_11b="PTB Chimpanzee Male BAC Library"  
BASE COUNT 526 a 371 c 199 g 39 t 65 others  
ORIGIN  
Query Match 15.0%; Score 23; DB 12; Length 1200;  
Best Local Similarity 100.0%; Pred. No. 0.22;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 50 CCAAAAACCAACGCAACAG 72  
|||||  
Db 254 CCAAAAACCAACGCAACAG 276  
|||||  
RESULT 13  
LOCUS BGS58209 1312 bp mRNA linear EST 29-MAY-2001  
DEFINITION 1034056008.xl C. reinhardtii CC-1690, normalized, Lambda Zap II  
ACCESSION BGS58209  
VERSION BGS58209.1 GI:14239393  
KEYWORDS EST.  
SOURCE Chlamydomonas reinhardtii.  
ORGANISM Chlamydomonas reinhardtii  
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.  
REFERENCE 1 (bases 1 to 1312)  
AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Lefebvre, P., Mcdermott, J.P., Sillflow, C., Stern, D. and Surzycki, R.  
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2  
JOURNAL Unpublished (2000)  
COMMENT Contact: Charles Hauser  
DCMB Box 91000  
Duke University  
Durham, NC 27708-1000

FEATURES  
source  
Tel: 919 613 8159  
Fax: 919 613 8177  
Email: chausere@duke.edu.  
Location/Qualifiers  
1. 1312  
/organism="Chlamydomonas reinhardtii"  
/strain="CC-1690 wild type mt+ 21gr"  
/db\_xref="taxon:3055"  
/clone\_11b="C. reinhardtii CC-1690, normalized, Lambda Zap II"  
/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey Mcdermott, combines cDNAs from CC-1690 cells grown to mid-log phase in YAP (acetate-containing) medium in the light, YAP medium in the dark, HS (minimal) medium in ambient levels of CO2 and HS medium bubbled with 5% CO2. Poly A RNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into Lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the Lambda Zap clones by superinfection with Exsist 4 (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."  
BASE COUNT 529 a 368 c 380 g 15 t 20 others  
ORIGIN  
Query Match 14.4%; Score 22; DB 10; Length 1312;  
Best Local Similarity 100.0%; Pred. No. 0.71;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 51 CAAAACCAACGCAACAG 72  
|||||  
Db 319 CAAAACCAACGCAACAG 340  
|||||  
RESULT 14  
LOCUS BF147109/c 217 bp mRNA linear EST 29-DEC-2000  
DEFINITION u070412.y1 McCarrey Eddy spermatocytes Mus musculus cDNA clone IMAGE:3657310 5' similar to SM:TCR2.MOUSE P11983 T-COMPLEX PROTEIN 1, ALPHA SUBUNIT B ; , mRNA sequence.  
ACCESSION BF147109  
VERSION BF147109.1 GI:11028504  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 217)  
AUTHORS Marra, W., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, F., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schunk, R., Ritter, E., Kohn, S., Shin, T., Jackson, F., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
TITLE The WashU-NCI Mouse EST Project 1999  
JOURNAL Unpublished (1999)  
COMMENT Contact: Marra W/WashU-NCI Mouse EST Project 1999  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:1419614  
Seq primer: Primer name ambiguous  
High quality sequence stop: 74.  
Location/Qualifiers  
1. 217  
/organism="Mus musculus"  
/strain="CD-1"  
/db\_xref="taxon:10090"

```
/clone="IMAGE:3657310"
/clone_lib="McCarrey Eddy spermatocytes"
/sex="male"
/tissue_type="spermatocytes, pooled from multiple mice"
/dev_stage="60 day"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: testis; Vector: pBluescript SK+ (Stratagene
); Site_1: XhoI; Site_2: EcoRI; cDNA oligo dT-primed
[5'-(GA)10-ACTAGCTCTGAGTCTTTTCTT-3'] and directionally
cloned using 5' linkers 5'-AATTGGCAGCAG-3' and
5'-CTCGTCCG-3'. Size selection of >400bp material gives
average insert size ranging from 1-2 kb. Library was mass
excised (from lambda-UniZAP-XR) and resulting
single-stranded phagemids were prepped and transformed
into DH10B. Library contains 98% recombinants.
References: J. Androl. 20:635-639 and Gene 25:263-269.
Library constructed and donated by J. McCarrey, Ph.D.
(Southwest Foundation for Biomedical Research, Dept. of
Genetics); excision done by E.M. Eddy, Ph.D. (National
Institutes of Health, National Institute of Environmental
Health Sciences). Original lambda-based library is
available through ATCC, catalog #63422."
```

BASE COUNT 46 a 38 c 60 g 73 t

ORIGIN

Query Match 12.4%; Score 19; DB 10; Length 217;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 TAGCAGAGATCTTGCAAAA 126  
|||||  
Db 193 TAGCAGAGATCTTGCAAAA 175

## RESULT 15

CNS02EDQ 404 bp DNA linear GSS 13-MAY-2000  
LOCUS Tetradon nigroviridis genome survey sequence PUC-ori end of clone

DEFINITION 262M09 of library G from Tetradon nigroviridis, genomic survey  
sequence.

ACCESSION AL193607.1 GI:7831713

VERSION GSS: genome survey sequence.

KEYWORDS Tetradon nigroviridis.

SOURCE Tetradon nigroviridis

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Percomorphna; Tetraodontiformes;

Tetraodontidae; Tetradon.

1 (bases 1 to 404)  
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,

Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and

Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetradon nigroviridis

Unpublished  
2 (bases 1 to 404)

Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,

Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,

Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using

Tetradon nigroviridis DNA sequence

Unpublished  
3 (bases 1 to 404)

Genoscope.  
Direct Submission

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

This sequence is a single read and was generated as part of a large

scale clone-end sequencing project of the Tetradon nigroviridis

genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetradon>.

Location/Qualifiers

source 1..404  
/organism="Tetradon nigroviridis"  
/db\_xref="taxon:99883"  
/clone="262M09"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C0AG262AG05SP1-end ;  
PUC-ori"

BASE COUNT 103 a 104 c 102 g 84 t 11 others

ORIGIN

Query Match 12.4%; Score 19; DB 12; Length 404;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 AAAAACCACAGCAAAACA 41  
|||||  
Db 352 AAAAACCACAGCAAAACA 370

Search completed: August 2, 2002, 22:41:33  
Job time: 30148 sec

## FEATURES

Mon Aug 5 11:51:48 2002

us-08-973-363-2.011.rst

Page 9



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 01:42:15 ; Search time 10310.5 Seconds  
(without alignments) 310.532 Million cell updates/sec

Title: US-08-973-363-3

Sequence: 1 ATTTCCTGATGATCGAGA.....CACAAAGCTTCGTGTGCA 153

Scoring table:

OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ha.\*  
2: gb\_hng.\*  
3: gb\_in.\*  
4: gb\_on.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_on.\*  
21: em\_of.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*  
29: em\_vl.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_in.\*  
32: em\_htg\_other.\*  
33: em\_htgo\_in.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No. Query Match Length DB ID Description

Result	Query	Match	Length	DB	ID	Description
1	103	67.3	6508	6	A58684	A58684 Sequence 3
2	103	67.3	6508	6	A58684	A58684 Sequence 10
3	103	67.3	6508	6	A58684	A58684 Sequence 10
4	64	41.8	2292	5	D14316	D14316 delta-eryt
5	47	30.7	153	6	A58685	A58685 Sequence 4
6	38	24.8	153	6	A58686	A58686 Sequence 15
7	25	16.3	1511	6	A58683	A58683 Sequence 2
8	25	16.3	5349	10	AR029026	AR029026 Sequence
9	25	16.3	5349	10	MUSCHD1X	MUSCHD1X
10	23	15.0	5947	9	AF06513	AF06513 Mouse DNA-b
11	23	15.0	101220	9	AC092372	AC092372 Homo sapi
12	23	15.0	134365	9	AC012624	AC012624 Homo sapi
13	23	15.0	143079	9	AC021449	AC021449 Homo sapi
14	23	15.0	143079	9	AC008531	AC008531 Homo sapi
15	23	15.0	193446	9	AC091946	AC091946 Homo sapi
16	23	15.0	219258	9	AC022121	AC022121 Homo sapi
17	23	15.0	276181	2	AC092382	AC092382 Homo sapi
18	21	13.7	92881	2	AC108265	AC108265 Rattus no
19	21	13.7	92881	2	AF181824	AF181824 Aegolius
20	20	13.1	2754	5	AF181824	AF181824 Nymphicus
21	20	13.1	2754	5	AF181824	AF181824 Gallus ga
22	20	13.1	2943	5	AF181824	AF181824 Gallus ga
23	20	13.1	147080	9	AL391666	AL391666 Human DNA
24	20	13.1	147891	2	AL365191	AL365191 Homo sapi
25	20	13.1	183892	2	AC019074	AC019074 Homo sapi
26	20	13.1	185204	2	AL446026	AL446026 Homo sapi
27	20	13.1	302097	2	AC098997	AC098997 Rattus no
28	19	12.4	3347	3	DM414676	DM414676 Drosophila
29	19	12.4	3347	3	DM414676	DM414676 Drosophila
30	19	12.4	30636	3	CEC31910	CEC31910 Caenorhabdit
31	19	12.4	51832	2	AC101875	AC101875 Mus muscu
32	19	12.4	51832	2	AC101875	AC101875 Mus muscu
33	19	12.4	110000	2	LMF18K18.01	LMF18K18.01
34	19	12.4	123527	9	AL353133	AL353133 Human DNA
35	19	12.4	127381	9	AL353133	AL353133 Human DNA
36	19	12.4	129473	9	AC040158	AC040158 Homo sapi
37	19	12.4	130138	9	AB053170	AB053170 Homo sapi
38	19	12.4	136500	9	AC097067	AC097067 Homo sapi
39	19	12.4	143900	2	AC092137	AC092137 Homo sapi
40	19	12.4	157600	2	AC092539	AC092539 Homo sapi
41	19	12.4	165536	9	AC093600	AC093600 Homo sapi
42	19	12.4	169514	9	AC064869	AC064869 Homo sapi
43	19	12.4	179619	2	AC018596	AC018596 Homo sapi
44	19	12.4	182199	2	AC021805	AC021805 Homo sapi
45	19	12.4	182500	2	AC026988	AC026988 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS A58684  
DEFINITION Sequence 3 from Patent WO9639505.  
ACCESSION A58684  
VERSION A58684.1 GI:3714247  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 to 153  
AUTHORS Griffiths,R. and Tjandi,R.  
TITLE AYIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
COMMENT ISTS INNOVATION (GB) 5906996 961224.  
FEATURES  
source  
1..153  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 58 a 40 c 31 g 24 t  
ORIGIN

Query Match 100.0%; Score 153; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 4.3e-78;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTTACCTGATATCCAGCAAGAAACCCAGCAAGAGCTACAGACCAAGAAACC 60  
|||||  
DB 1 ATTTTACCTGATATCCAGCAAGAAACCCAGCAAGAGCTACAGACCAAGAAACC 60

QY 61 CAGGCAAGAGCTACAGAGCCGTCGAGACTACTCTATTAAATTACGATAAGACCTT 120  
|||||  
DB 61 CAGGCAAGAGCTACAGAGCCGTCGAGACTACTCTATTAAATTACGATAAGACCTT 120

QY 121 GCAAGAAAGGAGACACAAAGGCTTGCTGTGCA 153  
|||||  
DB 121 GCAAGAAAGGAGACACAAAGGCTTGCTGTGCA 153

RESULT 2  
A58691 6608 bp DNA linear PAT 06-MAR-1998  
LOCUS Sequence 10 from Patent W09639505.  
ACCESSION A58691  
VERSION A58691.1 GI:3714250  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 6608)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 10 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
source 1. 6608  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN

Query Match 67.3%; Score 103; DB 6; Length 6608;  
Best Local Similarity 100.0%; Pred. No. 5.3e-49;  
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 CAAGAAACCCAGCAAGAGCTACAGAGCCGTCGAGACTACTCTATTAAATTACTGAA 110  
|||||  
DB 4100 CAAGAAACCCAGCAAGAGCTACAGAGCCGTCGAGACTACTCTATTAAATTACTGAA 4159

QY 111 TAAAGACCTTGCAAGAAAGAGACACAAAGGCTTGCGTGCA 153  
|||||  
DB 4160 TAAAGACCTTGCAAGAAAGAGACACAAAGGCTTGCGTGCA 4202

RESULT 3  
AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
LOCUS Gallus gallus chromo-helicase-DNA-binding on the Z chromosome  
DEFINITION protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete cds.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 6872)  
AUTHORS Griffiths,R. and Korn,R.M.

TITLE A CHD1 gene is z chromosome linked in the chicken Gallus domesticus  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
REFERENCE 2 (bases 1 to 6872)  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University, Glasgow G12 800, UK

FEATURES  
source 1. 6872  
Location/Qualifiers  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1. 6872  
/gene="CHD-Z"  
228. .5654  
/gene="CHD-Z"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the z chromosome protein"  
/protein\_id="AAC60282.1"  
/db\_xref="GI:2501846"  
/translation="MNGHSDSESVNRSGSSRSDDSGSAGSGSGSSGSSGSSGSS  
SOGSSDSESGSGSSESDTSERKQYAKPPKADSEFMKSPSILAVQSAV  
LKKQOQOQKAASDSESDSSSDSSSEDSSEETKKKKHKDEDMQSSGSGSGS  
DSEADGDSSCEESSEDEYEPKKNKYSKPPSRILKPKSGKSTQKROKQSDSE  
DDDEDYDKGSRQATVNVSYKEAEFEETKSDSDLEVCGEDVPTQDEDEFEETK  
SRIGKRGATGASTIYAVADDDPNAPFKSKSELEIOTLWKMKSHHNMHEET  
LKOQVKGANKLDNTKKQDETNRMLKNSPDVETYNQDELTDLHKQIVERTIL  
AHSNOKSAAGYDYYCKWQGLPYSECSWDGALAKKQARIDEFYFNRNOSTYPPK  
CIVLKQRPREFVALKQPSYIGGHESLELDYOLNGLNMLAHSGKNSCILLDEGLG  
KTIQITISFLNLYFHEHOLYGPFLRLPLSTLSQREIQTAPQONNAVYLDITSRN  
MRTHEMHMPOTKRLKFNILLTYEILKDKSEFLGLMNAFVIGDEARHLKDDSLY  
RLIDFKSMHRLITGTPLONSIKELKELTIMPEKSSMEDEFEHGGKSEYKAS  
LKELEPLILRRYKAKOVEKSLPAKBOILRMMSALQOYITWILTRINKALSKSSG  
STSGFLNIMMELKCCNHCYLIRPPDNEFYKKQALQHLIRSSGLILLDLILRLR  
EGRNRYLIFSQVVRMLDILAEYIKYROPFQRLDOSIKELKQALDHFNAGSDFC  
FLLSRAGGLGINLASADTVVLEFSDMNPQNDLOQARAHRIQKQOVNYYLTVKGS  
VEDILIERAKKKKMLDHLVYIORMDTGKVLHTGSPSSPPEEESAILFGAE  
LKEPGEFOEPQEMDIDELIKRAETREPERPLVGDILLQFVAFVANSNDDIE  
LEPERNSRMEETIPESORRIIEBEROKELEIYMLPMRCAKOISNGSEGRSR  
SRRYSGSDSDSITERRRPKRGRPTIPRENTKFSDAITRPIKSYKFGGLPRLRD  
AVARDAELDKSETDLRLGELVHNGCIKALDNSSGQERAGRGKVGPPFRISGV  
QVNAKLVISHHEELAPLHKISIPDEPERKYVPCHTKAHEDIDMGKEDSNLLVGI  
YEGYGSWEMIKMDPLSLTOKILDDPPKKQAOLOTRADYLLIKLKLDLAREAO  
RLAGAGNSKRRTKRNKKAKSIKEETKSSPOKSESPDEEDENKDEIVSVK  
HLHKITTEKNEDEKPEDDIGIKKAEERREKREKNEKELREKKEKDKKLEKED  
NKEKREKVKESYOKEREVEKVENMEKSENKESKIPDLDTPVHITATSPVPISE  
ESEELHOKTFEYVCKERMPPYKALQOLDBPERGLESERDLEHTROCLIKIGHITEC  
KEYTNEQIKQWRKMLMIFVSKFTEFDARKLKLKYLKALKROESQOHNQDNISNV  
THVIRNPDEVERLKEETTHNDSSSDSYSPRLSHOYDHKKDHOGQAYKKSPRRY  
SAPNSKHDRDHDKODSRVYSDSKNHRDLHRSDHNSNLEGLNKSRSRSHRS  
SPHRHSHDRSTSEYSHKSSRDYRHSQMDHRAHSGGSPSPILDQRPYSRSPDLG  
HRSPEHSSDHRSTPEHTWSSRKT"  
4341. .4604  
/gene="CHD-Z"  
/note="short insert found in longer variant mRNA of CHD-Z"

BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN

Query Match 67.3%; Score 103; DB 5; Length 6872;  
Best Local Similarity 100.0%; Pred. No. 5.2e-49;  
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 CAAGAAACCCAGCAAGAGCTACAGAGCCGTCGAGACTACTCTATTAAATTACTGAA 110  
|||||  
DB 4100 CAAGAAACCCAGCAAGAGCTACAGAGCCGTCGAGACTACTCTATTAAATTACTGAA 4159

QY 111 TAAAGACCTTGCAAGAAAGAGACACAAAGGCTTGCTGTGCA 153  
|||||  
DB 4160 TAAAGACCTTGCAAGAAAGAGACACAAAGGCTTGCTGTGCA 4202

RESULT 4  
LOCUS D14316 2292 bp mRNA linear VRT 03-FEB-1999  
DEFINITION delta-crystallin enhancer binding protein, complete sequence.  
ACCESSION D14316  
VERSION D14316.1 GI:391639  
KEYWORDS ORF2  
SOURCE Gallus gallus (Library: lambda gtl1) 13 day embryo lens cDNA to mRNA, clone JF11.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
REFERENCE 1 (bases 1 to 2292)  
AUTHORS Phasianidae, Gallus.  
TITLE Phasianidae, Gallus.  
JOURNAL Direct Submission  
SUBMITTED (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel. Aging and Cancer, Tohoku Univ.; 4-1 Seiryomachi, Aoba-ku, Sendai 980-77, Japan (Tel:022-272-9499, Fax:022-272-3982)  
2 (sites)  
REFERENCE 2 (bases 1 to 2292)  
AUTHORS Funahashi, J., Sekido, R., Mura, K., Kamachi, Y. and Kondoh, H.  
TITLE Delta-crystallin enhancer binding protein delta EPI is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis  
JOURNAL Development 119 (2), 433-446 (1993)  
MEDLINE 9411644  
REFERENCE 3 (bases 1 to 2292)  
AUTHORS Funahashi, J.  
JOURNAL Unpublished (1994)  
FEATURES  
source location/Qualifiers  
1..2292  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/tissue="lens"  
/clone\_lib="lambda gtl1"  
/dev\_stage="13 day embryo"  
257..1939  
/note="beta-gal fusion protein binds to blocks 10 to 3 of HN fragment of delta-crystallin enhancer."  
/codon\_start=1  
/product="ORF2"  
/protein\_id="BA03262.1"  
/db\_xref="GI:391640"  
/translation="MELKRCNHCYLKPPDNEFYNNQALQHLINSSGLLILDLKL LIRKRNRYLIFSQVPMDCITAEYLRQEPQRLDGSIKELRKQALDHPNAEG SEDPFLILNAGGLGINLASADTVIVIPSDNMPQDLQAAARHARGQKQVNYRL VYKGSVEFDILERAKKRVLDHLVYQRMDTGKTVLHSGSPSSSTPFNKELSAILK EDAELFKEPGEGEOPEDMDIDELKRAETRNENGPETVYMLPRMCAKOISFNGSE EDDILEPERSNNMEIIPESQRIIEEERKLELETFDAIRFIRKISYKKGCP LRRDAVRADELVDSETDLNRKGLVNHGCIKALKNSSGQRAGRIGVKGPTF RISGVYNAKIVISHEELAPLHKISIPDEEKKRYVPCRTKAHFDIDWCKEDDSN LLVGIYEYAVGSMQCVNCSRESQNTLEILDDPDRTPRONSRYRVOTTSINY"  
BASE COUNT 827 a 385 c 546 g 534 t  
ORIGIN

Query Match 41.8%; Score 64; DB 5; Length 2292;  
Best Local Similarity 100.0%; Pred. No. 3.1e-26;  
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 CAGCTACGAGACCGTGCACATCTATTAAATGATGAAAGCCTGCAGAAAG 129  
|||||  
DB 1900 CAGCTACGAGACCGTGCACATCTATTAAATGATGAAAGCCTGCAGAAAG 1959  
QY 130 GAAG 133  
|||||  
DB 1960 GAAG 1963  
RESULT 5

A58685  
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 4 from Patent W09639505.  
ACCESSION A58685  
VERSION A58685.1 GI:3714248  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths, R. and Tlward, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES  
source Location/Qualifiers  
1..153  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 56 a 36 c 31 g 30 t  
ORIGIN

Query Match 30.7%; Score 47; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 2.9e-16;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACCTCATTAATTTACTGATTAAGACCTTCGAGAAAGAACGACA 137  
|||||  
DB 91 TACCTCATTAATTTACTGATTAAGACCTTCGAGAAAGAACGACA 137

RESULT 6  
LOCUS A58696 1316 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 15 from Patent W09639505.  
ACCESSION A58696  
VERSION A58696.1 GI:3714253  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1316)  
AUTHORS Griffiths, R. and Tlward, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 15 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES  
source Location/Qualifiers  
1..1316  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 493 a 205 c 308 g 304 t 6 others  
ORIGIN

Query Match 30.7%; Score 47; DB 6; Length 1316;  
Best Local Similarity 100.0%; Pred. No. 2.6e-16;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACCTCATTAATTTACTGATTAAGACCTTCGAGAAAGAACGACA 137  
|||||  
DB 1198 TACCTCATTAATTTACTGATTAAGACCTTCGAGAAAGAACGACA 1244

RESULT 7  
LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 5 from Patent W09639505.  
ACCESSION A58686  
VERSION A58686.1 GI:3714249

KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
source 1..153  
/db\_xref="taxon:32644"  
/organism="unidentified"  
BASE COUNT 60 a 34 c 31 g 28 t  
ORIGIN

Query Match 24.8%; Score 38; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 5.2e-11;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACCTCATTAATTACGATTAAGACCTTGCAAGAA 128  
Db 91 TACCTCATTAATTACGATTAAGACCTTGCAAGAA 128

RESULT 8  
A58683  
LOCUS A58683 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 2 from Patent WO9639505.  
ACCESSION A58683  
VERSION A58683.1 GI:3714246  
KEYWORDS unidentified.  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 2 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
source 1..153  
/db\_xref="taxon:32644"  
/organism="unidentified"  
BASE COUNT 61 a 37 c 27 g 28 t  
ORIGIN

Query Match 16.3%; Score 25; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 0.002;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCGTCGACACTACCTCAT 98  
Db 74 TACAGACCGTCGACACTACCTCAT 98

RESULT 9  
AR029026  
LOCUS AR029026 1311 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 3 from patent US 5859183.  
ACCESSION AR029026  
VERSION AR029026.1 GI:5940999  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unidentified.  
REFERENCE 1 (bases 1 to 1311)

AUTHORS de Lange,T., Steensel,Bvan. and Bianchi,A.  
TITLE Altered telomere repeat binding factor  
JOURNAL Patent: US 5859183-A 3 12-JAN-1999;  
FEATURES Location/Qualifiers  
source 1..1311  
/organism="unknown"  
BASE COUNT 441 a 257 c 316 g 297 t  
ORIGIN

Query Match 16.3%; Score 25; DB 6; Length 1311;  
Best Local Similarity 100.0%; Pred. No. 0.0018;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TACAGACCGTCGACACTACCTCAT 98  
Db 1280 TACAGACCGTCGACACTACCTCAT 1304

RESULT 10  
MUSCHD1X  
LOCUS MUSCHD1X 5349 bp mRNA linear ROD 19-SEP-1996  
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.  
ACCESSION L10410.1  
VERSION L10410.1 GI:455014  
KEYWORDS DNA binding protein.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 5349)  
AUTHORS Delmas,Y., Stokes,D.G. and Perry,R.P.  
TITLE A mammalian DNA-binding protein that contains a chromodomain and an SNE2/SWI2-like helicase domain  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
MEDLINE 93211972  
REFERENCE 2 (bases 1 to 5349)  
AUTHORS Perry,R.P.  
TITLE Direct Submission  
JOURNAL Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, PA 19111, USA  
COMMENT On Feb 16, 1994 this sequence version replaced gi:293322.  
FEATURES Location/Qualifiers  
source 1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/product="DNA-binding protein"  
/protein\_id="AA08486.1"  
/db\_xref="GI:455015"  
/translation="MNGHSDSESVRNGSGESSQSGDCCSAGSGSGSSGSSSDGSS  
SOSGSDSDSGSDSGSSESDTSRENVQAKPPYVDASFWSKSPSTLAQVORAML  
RKPOOAOOORPASSNSGSESDSDSSGAKRKHNDEDMQSSGSPSLQ  
SPSESEERDKSSCDGETSDYEPKRVRSRKONRSGKGIIGOKRROIDSEDE  
DDEDYDNRKSSRROATVAVSYKREDEKMTSDDLLEYGGEVYPQDEFEFETRYM  
DCRVGRKATGATTTTIVAEADGDPVAGFERKKEPEDIDYLLIKMGWSEHINWTEE  
TLQONVRGMRKLDNYKKRDKQETKRLKNAKSPEDVEYVQCQELDDLRKQIVERI  
IAHSNOKSAAGLPDYCKWQGLPYSECSMEDALISKFCOTCIDEFYSNQSQTIPFK  
DCKVLKORPRFVALKROPSTYIGHGHELELDYQINGLNLMLASHMCKGNSCILADEMG  
KTIQTIQSFNLNLFHEHOLYGPPLVYPTSLTSMQREIQTWASOMNAVYVYGDINSR  
NMRTHEWMPQTKRLKFNILITTEILIKDRAFLGGLMWATGVDEARLKNDSLSL  
YKTLIDFKSNHRLITGPELQNSLEKMLSLHIFIMPEKSSWEDEEHEGKREGEYA  
SLHKELEPFLRRVYKDYKSLPAKVEQILIRMEMSALQKQYKWLITRNKALISGSK  
GSGSEPLINMELEKCCNCHYLIRPDNNEFNRKQALQHLIRSSGKLTLLDKLIRL  
RGRGNRVILFESQMVRLDITLAIYKRYROPPORLDGSLKGLRKQALHFNESGSEDF  
CFLSTRAGGLGILNASADTVYIPSDNRPONDLOAQRARHIGQKQKQYNYRIVYTKG  
SVEEDILERRAKRMVLDHVIQRMQDTGTVLHTGSAPSSSTPNNKEELSAITLKGA  
ELFKEPEGEQEPQEMDIDELIKRAETHENEGLPSVGDLELLSQKRVANFSNMDDDI



[illegible]

Query Match	Best Local Similarity	100.0%	Score 23	DB 9	Length 5947
Matches 23	Conservative	0	Mismatches	0	Indels 0
Gaps	0				
DB 4070	CAGACCGCTGACACTACCTCAT	4092			
RESULT 12					
AC092372/c					
LOCUS	AC092372	101220 bp	DNA	linear	PRI 07-DEC-2001
DEFINITION	Human sapiens chromosome 5 clone RP11-58M12, complete sequence.				
ACCESSION	AC092372				
VERSION	AC092372.3	GI:17402768			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
TITLE	1 (bases 1 to 101220)				
REFERENCE	DOE Joint Genome Institute and Stanford Human Genome Center.				
AUTHORS	Direct Submission				
TITLE	2 (bases 1 to 101220)				
REFERENCE	DOE Joint Genome Institute.				
AUTHORS	Direct Submission				
TITLE	Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint				
REFERENCE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
AUTHORS	3 (bases 1 to 101220)				
TITLE	DOE Joint Genome Institute and Stanford Human Genome Center.				
JOURNAL	Direct Submission				
TITLE	Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell				
REFERENCE	Drive, Walnut Creek, CA 94598, USA				
AUTHORS	On Dec 7, 2001 this sequence version replaced gi:15290448.				
JOURNAL	Draft Sequence Produced by DOE Joint Genome Institute				
COMMENT	www.jgi.doe.gov				
	Finishing Completed at Stanford Human Genome Center				
	www.sbgc.stanford.edu				
	Quality: Phrap Quality >=40 100% of Sequence;				
	Estimated Total Number of Errors is 0.				
	NOTE: This insert is not the entire sequence of the clone (entire				
	sequence is 146,7kb) It is clipped at the overlap with AC012624.				
	The number of bases overlapped is 90404.				
FEATURES	Location/Qualifiers				
source	1..101220				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="5"				
	/clone="RP11-58M12"				
BASE COUNT	34122 a 18862 c 17827 g 30409 t				

```

ORIGIN

Query Match      15.0%; Score 23; DB 9; Length 101220;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 76 CAGACCCGTGACAGACTACTCAT 98
|||||
Db 26887 CAGACCCGTGACAGACTACTCAT 26865

RESULT 13
AC012624 134365 bp DNA linear PRI 21-JUL-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
DEFINITION AC012624
ACCESSION AC012624.6 GI:14993679
VERSION HTG.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 134365)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 134365)
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
source
1. 134365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN

Query Match      15.0%; Score 23; DB 9; Length 134365;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 76 CAGACCCGTGACAGACTACTCAT 98
|||||
Db 118292 CAGACCCGTGACAGACTACTCAT 118314

RESULT 14
AC021449 143079 bp DNA linear HTG 10-SEP-2000
LOCUS Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION AC021449
VERSION AC021449.3 GI:10047806
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE 1 (bases 1 to 143079)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-58M12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 143079)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Deaileano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karats,A., Klein,J.,
Landers,T., Lebecky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Maddonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,D.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,Y., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,K.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2000 this sequence version replaced gi:1407963.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html
-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L5154
Center clone name: 58_M_12
-----
Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 134743 bases at least Q40
Consensus quality: 139227 bases at least Q30
Consensus quality: 140814 bases at least Q20
Insert size: 14400; agarose-fp
Insert size: 142179; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.7 in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 38820: contig of 38820 bp in length
* 38821 38920: gap of 100 bp
* 38921 40411: contig of 1491 bp in length
* 40412 40511: gap of 100 bp
* 40512 43279: contig of 2768 bp in length
* 43280 43379: gap of 100 bp
* 43380 46905: contig of 3526 bp in length
* 46906 47005: gap of 100 bp
* 47006 51830: contig of 4825 bp in length
* 51831 51930: gap of 100 bp
* 51931 62619: contig of 10689 bp in length
* 62620 62719: gap of 100 bp
* 62720 75408: contig of 12689 bp in length
* 75409 75508: gap of 100 bp
* 75509 92516: contig of 17008 bp in length

```

```

* 92517 92616: gap of 100 bp
* 92617 106409: contig of 13793 bp in length
* 106410 106509: gap of 100 bp
* 106510 143079: contig of 36570 bp in length.
Location/Qualifiers
1. 143079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-58M12"
/clone.lib="RP11-58M12"

```

```

misc_feature
1. 38820
/feature="assembly_fragment"
clone_end:SP6
vector_side:left"
38921..40411
/misc_feature
/feature="assembly_fragment"
40312..43279
/misc_feature
/feature="assembly_fragment"
43380..46905
/misc_feature
/feature="assembly_fragment"
47006..51830
/misc_feature
/feature="assembly_fragment"
51931..62619
/misc_feature
/feature="assembly_fragment"
62720..75408
/misc_feature
/feature="assembly_fragment"
75509..92516
/misc_feature
/feature="assembly_fragment"
92617..106409
/misc_feature
/feature="assembly_fragment"
106510..143079
/feature="assembly_fragment"
clone_end:77
vector_side:right"
BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others
ORIGIN

```

```

Query Match 15.0% Score 23; DB 2; Length 143079;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 76 CAGACCCGTCGACACTACTCAT 98
|||||
Db 116192 CAGACCCGTCGACACTACTCAT 116214

```

```

RESULT 15
AC008531 145659 bp DNA linear HTG 14-FEB-2001
LOCUS Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,
DEFINITION 7 ordered pieces.
AC008531
AC008531.3 GI:12830078
VERSION HTG: HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 145659)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145659)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

```

```

-----
Project Information
Center Project Name: 369535
Center clone name: CIT-HSPC_480B11

```

```

Summary Statistics
Consensus quality: 139128 bases at least Q40
Consensus quality: 142556 bases at least Q30
Consensus quality: 143744 bases at least Q20
Estimated insert size: 148000: pulse-field gel estimation
Estimated insert size: 145059: sum-of-contigs estimation
Quality coverage: 6.27 in Q20 bases; pulse-field gel estimation
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
consists of 7 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 56174: contig of 56174 bp in length
56175 56274: gap of unknown length
56275 100874: contig of 44600 bp in length
100875 100974: gap of unknown length
100975 113127: contig of 12153 bp in length
113128 113227: gap of unknown length
113228 118190: contig of 4963 bp in length
118191 118290: gap of unknown length
118291 119694: contig of 1404 bp in length
119695 119794: gap of unknown length
119795 123297: contig of 3503 bp in length
123298 123397: gap of unknown length
123398 145659: contig of 22262 bp in length.
Location/Qualifiers
1 145659
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-480B11"
/clone.lib="Caltech human BMC library C"

```

```

BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others
ORIGIN

```

```

Query Match 15.0% Score 23; DB 2; Length 145659;
Best Local Similarity 100.0%; Pred. No. 0.021;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 76 CAGACCCGTCGACACTACTCAT 98
|||||
Db 46815 CAGACCCGTCGACACTACTCAT 46837

```

```

Search completed: August 3, 2002, 01:42:44
Job time: 40734 sec

```

...

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 01:50:02 ; Search time 949.74 Seconds

Title: US-08-973-363-3

Sequence: 1 ATTTACCTGATGCCAGA.....CACAAAGGCTTGCTGGTGCA 153

Account	Balance
Gapop 60.0	Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

word size :

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum	DB seq	length:	2000000000
Minimum	DB seq	length:	0

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_032802:\*

1	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT*
2	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT*
3	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT*
4	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT*
5	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT*
6	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT*
7	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT*
8	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT*
9	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT*
10	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT*
11	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT*
12	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT*
13	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT*
14	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT*
15	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT*
16	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT*
17	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT*
18	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT*
19	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT*
20	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT*
21	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT*
22	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001.DAT*
23	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT*
24	/SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2003.DAT*

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	153	100.0	153	18	AA42757	Chick CHD-1A gene
2	103	67.3	6608	18	AA42751	Chicken CHD-1A ge
3	47	30.7	153	18	AA42758	Chick CHD-W gene
4	47	30.7	153	18	AA42754	Chick CHD-W gen
5	38	24.8	153	18	AA42759	Great tit CHD-W g
6	25	16.3	153	18	AA42756	Mouse CHD-1 gene
7	25	16.3	1311	19	AAV59280	Altered telomere
8	18	11.8	401	22	AA180568	Human polynucleot
9	18	11.8	552	22	AA568219	Human lung tumour

[illegible]

PI Griffiths R, Tiwari B;  
 XX  
 DR WPI: 1997-043127/04.  
 DR P-PSDB; AAW08147.  
 XX  
 XX Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determin. and to control sex of progeny  
 XX  
 PS Claim 8; Fig 3; 76pp; English.  
 CC Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology  
 CC to portions of the chicken CHD-1A (A = Avian) gene (AA142757),  
 CC chicken CHD-W (W refers to the W chromosome) gene (see also AA142758),  
 CC and the great tit CHD-W gene (see also AA142759). Translated amino  
 CC acid sequences of this region are provided in AAW08146-49. The  
 CC CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes  
 CC determine sex in birds and can be used to identify the sex of an  
 CC embryo, foetus etc. and to manipulate the sex of progeny.  
 XX  
 SQ Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

Query Match 100.0%; Score 153; DB 18; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-70;  
 Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AATTACTGATGATCCAGACAGAAACCAGCAAGCAGCTACAGACCAAGAAACC 60  
 DB 1 attttaccctgagatcccaagaagaaccacccagcaagcagctacagaccagaacc 60  
 OY 61 CAGGCAAGCAGCTACAGACCCGTCGAGACTACCTATTAAATTACTGAATAAGACTT 120  
 DB 61 cagcagaagcagctacagaccgctgcagactactcattactgaataagacctt 120  
 OY 121 GCAAGAAAGGAGACCAAGACCTTGCTGCTGCA 153  
 DB 121 gcaagaagaagacacaaagccttgctgtgtgca 153

RESULT 2  
 AA142751  
 ID AA142751 standard; cDNA; 6608 BP.  
 XX  
 AC AA142751;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chicken CHD-1A gene.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
 KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 XX Key Location/Qualifiers  
 FH CDS 228..5390  
 FT /\*tag= a  
 FT  
 FT  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PE 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI: 1997-043127/04.  
 DR P-PSDB; AAW08147.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 XX birds - used for sex determin. and to control sex of progeny

PT birds - used for sex determin. and to control sex of progeny  
 XX  
 XX Claim 1; Fig 5; 76pp; English.  
 PS  
 CC The chicken CHD-W gene (AA142754) acting alone or in conjunction with  
 CC the closely related CHD-1A gene (AA142751) is suggested to initiate  
 CC female development in birds. The sequence of CHD-1A was deduced  
 CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
 CC library using a great tit CHD-W sequence (see also AA142755) as probe.  
 CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
 CC gene (see also AA142756-57). It is located on an autosome or X  
 CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
 CC specific signal on hybridisation to genomic DNA of a non-ratite  
 CC bird and can be used for sex determin. of a bird. CHD-1A nucleic  
 CC acids can also be used to control the sex of the progeny of a bird.  
 XX  
 SQ Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

Query Match 67.3%; Score 103; DB 18; Length 6608;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-44;  
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 51 CAGAAACCCAGCAAGCAGCTACAGACCCGTCGAGACTACCTATTAAATTACTGAA 110  
 DB 4100 caagaaccccccagcaagcagctacagaccgctgcagactactcattactga 4159  
 OY 111 TAAAGACTTGCAAGAAAGACACCAAGCTTGCTGCTGCA 153  
 DB 4160 taagaccttgcaagaagaagcacaagaagccttgctgtgtgca 4202

RESULT 3  
 AA142758  
 ID AA142758 standard; DNA; 153 BP.  
 XX  
 AC AA142758;  
 XX  
 DT 12-MAR-1997 (first entry)  
 XX  
 DE Chick CHD-W gene fragment.  
 XX  
 KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
 KW CHD-1A; CHD-W; W chromosome; ss.  
 XX  
 OS Gallus sp.  
 XX  
 XX Key Location/Qualifiers  
 FH misc\_difference 52..81  
 FT /\*tag= a  
 FT /note= "bases 52-81 are a repeat of bases 22-51  
 FT and are ignored in the translated amino  
 FT acid sequence given in Fig 3"  
 FT  
 FT  
 PN WO9639505-A1.  
 XX  
 PD 12-DEC-1996.  
 XX  
 PE 05-JUN-1996; 96WO-GB01341.  
 XX  
 PR 06-JUN-1995; 95GB-0011439.  
 XX  
 PA (ISIS-) ISIS INNOVATION LTD.  
 XX  
 PI Griffiths R, Tiwari B;  
 XX  
 DR WPI: 1997-043127/04.  
 DR P-PSDB; AAW08148.  
 XX  
 PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 XX birds - used for sex determin. and to control sex of progeny

Claim 8; Fig 3; 76pp; English.

```

XX Bases 3855-3977 (AAAT42756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AAAT42757),
CC chicken CHD-W (W refers to the W chromosome) gene (AAAT42758) and
CC and the great tit CHD-W gene (AAAT42759). Translated amino acid
CC sequences of this region are provided in AAM08146-49. The CHD-1A
CC (see also AAAT42751) and CHD-W (see also AAAT42754-55) genes determine
CC sex in birds and can be used to identify the sex of an embryo,
CC foetus etc. and to manipulate the sex of progeny.
XX
SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other:

Query Match          30.7%; Score 47; DB 18; Length 153;
Best Local Similarity 100.0%; Pred. No. 7.1e-15;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 91 TACCTCATTAATTACTGATTAAGACCTTCGACAGAAAGACGACA 137
   ||||||||||||||||||||||||||||||||||||||||||||
DB 91 taccctcataattactgataaagaccttcgaagaagaagacaca 137

RESULT 4
AAAT42754
ID AAAT42754 standard; CDNA; 1316 BP.
XX
AC AAAT42754;
XX
DT 12-MAR-1997 (first entry)
XX
DE Chicken CHD-W gene (partial sequence).
XX
KM Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX
OS Gallus sp.
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PE 05-JUN-1996; 96MO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
DR WPI; 1997-043127/04..
XX
FT Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determ. and to control sex of progeny
XX
PS Claim 1; Fig 8; 76pp; English.
XX
CC The chicken CHD-W gene (AAAT42754) acting alone or in conjunction with
CC the closely related CHD-1A gene (AAAT42751) is suggested to initiate
CC female development in birds. The sequence of CHD-W was deduced
CC from 2 clones isolated from a 10-day chick embryo library using
CC a fragment of the CHD-1A gene as a probe. The CHD-W gene is
CC located on the W chromosome. Probes based on CHD-W and CHD-1A give
CC a W chromosome-specific signal on hybridisation to genomic DNA of a
CC non-rare bird and can be used for sex determin. of a bird. CHD-W
CC nucleic acids can also be used to control the sex of progeny of a
CC bird.
XX
SQ Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other:

Query Match          30.7%; Score 47; DB 18; Length 1316;
Best Local Similarity 100.0%; Pred. No. 7.3e-15;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 91 TACCTCATTAATTACTGATTAAGACCTTCGACAGAAAGACGACA 137
   ||||||||||||||||||||||||||||||||||||||||||||
DB 1198 taccctcataattactgataaagaccttcgaagaagaagacaca 1244

RESULT 5
AAAT42759
ID AAAT42759 standard; DNA; 153 BP.
XX
AC AAAT42759;
XX
DT 12-MAR-1997 (first entry)
XX
DE Great tit CHD-W gene fragment.
XX
KM Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KW CHD-1A; CHD-W; W chromosome; ss.
XX
OS Parus major.
XX
FH Key Location/Qualifiers
FT misc_difference 52..81
FT /*tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
PN WO9639505-A1.
XX
PD 12-DEC-1996.
XX
PE 05-JUN-1996; 96MO-GB01341.
XX
PR 06-JUN-1995; 95GB-0011439.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Griffiths R, Tiwari B;
XX
DR WPI; 1997-043127/04.
XX
DR P-PSDB; AAM08149.
XX
PT Avian chromodomain-helicase-DNA binding genes determine sex in
PT birds - used for sex determ. and to control sex of progeny
XX
PS Claim 8; Fig 3; 76pp; English.
XX
CC Bases 3855-3977 (AAAT42756) of the mouse CHD-1 gene show homology
CC to portions of the chicken CHD-1A (A = Avian) gene (AAAT42757),
CC chicken CHD-W (W refers to the W chromosome) gene (AAAT42758) and
CC and the great tit CHD-W gene (AAAT42759). Translated amino acid
CC sequences of this region are provided in AAM08146-49. The CHD-1A
CC (see also AAAT42751) and CHD-W (see also AAAT42754-55) genes determine
CC sex in birds and can be used to identify the sex of an embryo,
CC foetus etc. and to manipulate the sex of progeny.
XX
SQ Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other:

Query Match          24.8%; Score 38; DB 18; Length 153;
Best Local Similarity 100.0%; Pred. No. 3.4e-10;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 91 TACCTCATTAATTACTGATTAAGACCTTCGACAGAA 128
   ||||||||||||||||||||||||||||||||||||||||||||
DB 91 taccctcataattactgataaagaccttcgaagaaga 128

RESULT 6
AAAT42756
ID AAAT42756 standard; DNA; 153 BP.

```

```

AC  AA142756;
XX
XX  12-MAR-1997 (first entry)
DE  Mouse CHD-1 gene (bases 3855-9777).
XX
XX  Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KW  CHD-1; CHD-W; W chromosome; ss.
XX
XX  Mus sp.
OS
XX
XX  Key Location/Qualifiers
FH  misc_difference 52..81
FT  /tag= a
FT  /note= "bases 52-81 are a repeat of bases 22-51
FT  and are ignored in the translated amino
FT  acid sequence given in Fig 3"
XX
XX  W09639505-A1.
XX
XX  12-DEC-1996.
XX
XX  05-JUN-1996; 96WO-GB01341.
XX
XX  06-JUN-1995; 95GB-0011439.
XX
XX  (ISIS-) ISIS INNOVATION LTD.
XX
XX  Griffiths R, Tiwari B;
XX
XX  WPI; 1997-043127/04.
XX
XX  P-PSDB; AAW08146.
XX
XX  Avian chromodomain-helicase-DNA binding genes determine sex in
XX  birds - used for sex determ. and to control sex of progeny
XX
XX  Claim 8; Fig 3; 76pp; English.
XX
XX  Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX  to portions of the chicken CHD-1A (A = Avian) gene (see also
XX  and AA142757), chicken CHD-W (W refers to the W chromosome) gene
XX  (see also AA142758) and the great tit CHD-W gene (see also AA142759).
XX  Translated amino acid sequences of this region are provided in
XX  CC AAW08146-49. The CHD-1A (see also AA142751) and CHD-W (see also
XX  CC AA142754-55) genes determine sex in birds and can be used to identify
XX  the sex of an embryo, fetus etc. and to manipulate the sex of
XX  CC progeny.
XX
XX  Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;
XX
XX  Query Match 16.3%; Score 25; DB 18; Length 153;
XX  Best Local Similarity 100.0%; Pred. No. 0.002;
XX  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  74 TACAGACCGGTGACAGACTACTCAT 98
XX  ||||||||||||||||||||
XX  Db 74 tacagaccggtgcagactactcat 98
XX
XX  RESULT 7
XX  ID AAV59280 standard; cDNA; 1311 BP.
XX
XX  AAV59280;
XX
XX  14-DEC-1998 (first entry)
XX
XX  Altered telomere repeat binding factor 1 gene.
XX
XX  ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;
XX  telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.
XX

```

```

OS  Homo sapiens.
XX  Synthetic.
XX
XX  Key Location/Qualifiers
FH  CDS 1..1311
FT  /tag= a
FT  /product= "A-TRP"
XX
XX  W09836066-A1.
XX
XX  20-AUG-1998.
XX
XX  13-FEB-1998; 98WO-US02765.
XX
XX  04-FEB-1998; 98US-0018628.
XX  13-FEB-1997; 97US-0800264.
XX
XX  (UYRO ) UNIV ROCKEFELLER.
XX
XX  Bianchi A, De Lange T, Van Steensel B;
XX
XX  WPI; 1998-480769/41.
XX  P-PSDB; AAV59280.
XX
XX  Nucleic acid encoding altered telomere repeat binding protein and
XX  related vectors - transformants, hetero-dimers and antibodies, used
XX  to inhibit shortening of telomerases caused by ageing or disease,
XX  also used to extend life of cells in culture
XX
XX  Claim 14; Page 110-111; 163pp; English.
XX
XX  The altered vertebrate telomere repeat binding protein (A-TRF) has a
XX  telomere repeat binding factor (TRF) dimerisation domain, and forms a
XX  hetero-dimer with TRF, preventing it from binding to the specified repeat
XX  sequence. A-TRF, optionally expressed by gene therapy, is used to
XX  inhibit shortening of telomeres associated with ageing (for cosmetic
XX  purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,
XX  atrophy of the skin, age-related macular degeneration, atherosclerosis,
XX  tumours and viral (including human immune deficiency virus) infection.
XX  Cells expressing A-TRF also have an increased life span in vitro, e.g.
XX  for expression of recombinant proteins or where intended for subsequent
XX  transplant or for testing, eliminating the need for transformation.
XX
XX  Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;
XX
XX  Query Match 16.3%; Score 25; DB 19; Length 1311;
XX  Best Local Similarity 100.0%; Pred. No. 0.0021;
XX  Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  74 TACAGACCGGTGACAGACTACTCAT 98
XX  ||||||||||||||||||||
XX  Db 1280 tacagaccggtgcagactactcat 1304
XX
XX  RESULT 8
XX  ID AAI80568/c
XX  AAI80568 standard; cDNA; 401 BP.
XX
XX  AAI80568;
XX
XX  06-NOV-2001 (first entry)
XX
XX  Human polynucleotide SEQ ID NO 628.
XX
XX  Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX  vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX  tissue growth factor; immunomodulatory; cancer; leukaemia;
XX  nervous system disorders; arthritis; inflammation; ss.
XX
XX  Homo sapiens.
XX
XX  W0200164835-A2.
XX

```



```

XX 07-SEP-2001.
XX 26-FEB-2001; 2001WO-US04927.
XX 28-FEB-2000; 2000US-0515126.
XX 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEO INC.
XX Tang YF, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AA000637.
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukemia, inflammation and immune
XX disorders -
XX
XX Claim 1; SEQ ID NO 628; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX the encoded proteins (AA000010-AA013910) that exhibit activity relating to
XX cytokine cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoietic regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 401 BP; 113 A; 87 C; 108 G; 93 T; 0 other:
SQ
Query Match 11.8%; Score 18; DB 22; Length 401;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 CAGACAGAAACCCGAG 34
    |||||||||||||||
DB 304 CAGACAGAAACCCGAG 287

RESULT 9
ID AAF68219 standard; cDNA: 552 BP.
XX AAF68219;
AC AAF68219;
XX
XX 12-APR-2001 (first entry)
DE Human lung tumour protein related nucleotide sequence SEQ ID NO:137.
XX
XX Human; lung cancer; lung tumour; lung tumour protein; gene therapy;
XX lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;
XX cytostatic; antisense inhibition; ss.
XX
XX Homo sapiens.
XX
XX WO200100828-A2.
XX
XX 04-JAN-2001.
XX
XX 30-JUN-2000; 2000MO-US18061.
XX
XX 30-JUN-1999; 99US-0346492.
XX 15-OCT-1999; 99US-0419356.
XX 17-DEC-1999; 99US-0466867.

```

```

PR 30-DEC-1999; 99US-0476300.
PR 06-MAR-2000; 2000US-0519642.
PR 22-MAR-2000; 2000US-0533077.
PR 10-APR-2000; 2000US-0546259.
PR 27-APR-2000; 2000US-0560406.
XX 05-JUN-2000; 2000US-0589184.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;
XX Retter MW, Mannion J;
XX WPI; 2001-071488/08.
XX
XX Lung tumor-associated proteins and the nucleic acids that encode them,
XX useful for preventing, diagnosing and treating lung cancer -
XX
XX Example 1; Page 193; 436pp; English.
XX
XX The present invention describes immunogenic portions of lung tumour-
XX associated proteins (I) and the nucleic acids (NAs) that encode them.
XX (I) have cytostatic activity and can be used in gene therapy, antisense
XX inhibition and in vaccines. The NAs and the lung tumour-associated
XX proteins they encode may be used in the prevention, treatment and
XX diagnosis of diseases associated with their inappropriate expression,
XX especially lung cancers. For example, the NAs may be administered to
XX treat diseases by rectifying mutations or deletions in a patient's genome
XX that affect the activity of the protein by expressing inactive proteins
XX or to supplement the patient's own production of (I). Additionally, the
XX NAs may be used to produce the lung-tumour associated protein, according
XX to standard recombinant DNA methodology. Conversely, antisense NA
XX molecules may be administered to down regulate protein expression by
XX binding with the cells own genes and preventing their expression. The NA
XX and complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar NA sequences in
XX cancer. The (I) may be used as antigens in the production of antibodies
XX and in assays to identify modulators (agonists and antagonists) of the
XX expression and activity of the protein. AAF68083 to AAF68878 and
XX AAB76848 to AAB76878 represent human lung tumour protein related
XX nucleotide and protein sequences which are used in the exemplification
XX of the present invention.
XX
XX Sequence 552 BP; 220 A; 118 C; 69 G; 144 T; 1 other:
SQ
Query Match 11.8%; Score 18; DB 22; Length 552;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 97 ATTAATTCTGATATA 114
    |||||||||||||||
DB 492 attaatctactgataaa 509

RESULT 10
ID AAV49574 standard; cDNA to mRNA: 1409 BP.
XX AAV49574;
AC AAV49574;
XX
XX 21-OCT-1998 (first entry)
DE Human lymphoma cell line U937 clone HP10136 cDNA #1.
XX
XX Transmembrane domain; human; nutrition; cytokine; cell proliferation;
XX differentiation; immune system; stimulator; suppressor; regulator;
XX haematopoiesis; activin; inhibitor; chemotactic; chemokine; receptor;
XX haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
PH

```

FT	CDS	82..729
FT		/tag=a
FT		/product="transmembrane domain containing protein"
XX		
XX		
PN	W09821328-A2.	
PD	22-MAY-1998.	
XX		
PF	07-NOV-1997;	97WO-JP04056.
XX		
PR	13-NOV-1996;	96JP-0301429.
XX		
PA	(PROT-) PROTEGENE INC. (SAGA) SAGAMI CHEM RES CENTRE.	
XX		
PI	Kato S, Kobayashi M, Sekine S, Yamaguchi T;	
XX		
DR	WPI; 1998-297932/26.	
DR	P-PISDB; AAM64546.	
XX		
PT	Human protein having transmembrane domain - useful for, e.g. research and nutrition	
XX		
PS	Claim 4; Page 153-155; 205pp; English.	
CC	AADV9550-V49599 are cDNA sequences which encode human proteins containing a transmembrane domain. These proteins can be used for, e.g. research and nutrition, and may have cytokine and cell growth stimulating/suppressing, proliferation/differentiation, immune stimulating/suppressing, haematopoiesis regulating, tissue growth, activin/inhibin, chemotactic/chemokinetic, haemostatic and thrombolytic, receptor/ligand, anti-inflammatory or tumour inhibition activity.	
CC		
CC		
SO	Sequence 1409 BP; 377 A; 278 C; 325 G; 429 T; 0 other;	
Query Match                      11.8%; Score 18; DB 19; Length 1409; Best Local Similarity         100.0%; Pred. No. 9.2; Matches    18; Conservative    0; Mismatches    0; Indels    0; Gaps    0;		
QY	111 TAAAGACCTTGCAAGAAA 128 	
Db	842 taaagaccttgcagaaga 859	
RESULT 11		
ID	AAK56240	
XX	AAK56240 standard; CDNA; 1462 BP.	
AC	AAK56240;	
XX		
DT	16-JUL-1999 (first entry)	
DE	Human CBFBBA01 vesicle trafficking protein SEC22b gene.	
XX		
KW	Human; vesicle trafficking protein; SEC22b; CBFBBA01; cancer; autoimmune disease; diabetes mellitus; multiple sclerosis; diagnosis; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	W09921988-A1.	
PD	06-MAY-1999.	
XX		
PF	29-OCT-1997;	97WO-CN00115.
XX		
PR	29-OCT-1997;	97WO-CN00115.
XX		
PA	(UYSH-) UNIV SHANGHAI SECOND MEDICAL.	
XX		
PI	Wang Y, Wu J;	
XX		

DR		WPI: 1999-312961/26.
DR	P-PESD;	AAV09512.
XX	CBFBA01 protein useful in treating cancer, autoimmune disease,	
PT	diabetes mellitus and multiple sclerosis	
XX		
PS	Claim 2; Page 21; 29pp; English.	
XX		
CC	The present sequence encodes CBFBA01, which is a protein homologous to	
CC	murine vesicle trafficking protein sec22b. CBFBA01 is a	
CC	cytoplasmically-oriented integral membrane protein, located in the	
CC	endoplasmic reticulum and golgi membranes. CBFBA01 polynucleotide (I),	
CC	vectors containing (I) and recombinant host cells are useful for	
CC	recombinant production of CBFBA01. Both (I), CBFBA01 and antibodies	
CC	against CBFBA01 are useful as research reagents, for screening assays	
CC	and in diagnostic assays, especially for cancer, autoimmune disease,	
CC	diabetes mellitus and multiple sclerosis. Antagonists and agonists of	
CC	CBFBA01 can be used to inhibit or enhance, respectively, the activity	
CC	of CBFBA01 or expression of (I). Anti-CBFBA01 antibodies and CBFBA01	
CC	or its fragments can be used in vaccines.	
XX		
SQ	Sequence 1462 BP; 419 A; 286 C; 323 G; 434 T; 0 other:	
	Query Match                    11.8%; Score 18; DB 20; Length 1462;	
	Best Local Similarity       100.0%; Pred. No. 9.2;	
	Matches     18; Conservative     0; Mismatches     0; Indels     0; Gaps     0;	
OY	111 TAAAGACTTGCAAGAAA 128         Db     826 taaagaccttgcaaga aa 843	
RESULT 12		
ID	AAV20467/C	
XX	AAV20467 standard; DNA: 3225 BP.	
AC	AAV20467;	
XX		
DT	17-JUN-1998 (first entry)	
XX		
DE	Human c-myc oncogene.	
XX		
KM	Human; oncogene; proto-oncogene; neoplastic disease; anticancer;	
XX	cancer; antisense oligonucleotide; c-myc; ds.	
OS	Homo sapiens.	
XX		
PN	US5734039-A.	
XX		
PD	31-MAR-1998.	
XX		
PF	15-SEP-1994; 94US-0306691.	
XX		
PR	15-SEP-1994; 94US-0306691.	
PA	(UYUE-) UNIV JEFFERSON THOMAS.	
PI	Calabretta B, Skorski T;	
XX		
DR	WPI: 1998-229882/20.	
XX		
PT	Anticancer composition comprising two anti-sense oligo:nucleotide(s)	
XX	- targeting cytoplasmic and nuclear oncogene(s)	
PS	Claim 1; Column 131-134; 92pp; English.	
XX		
CC	The present sequence represents an oncogene from the present invention.	
CC	The present invention describes a composition which comprises two	
CC	antisense oligonucleotides. The first oligonucleotide is specific for a	
CC	cytoplasmic oncogene or proto-oncogene selected from ras, raf, Egr-1,	
CC	c-fos, c-src, c-kit, c-met, c-trk, c-src, c-abl, bcr-abl, c-fgr and	
CC	c-yes. The second oligonucleotide is specific for a nuclear oncogene or	

CC proto-oncogene selected from myc, jun, c-ets, c-fos, c-myc, B-myb,  
 CC c-rel, c-vav, c-ski, c-spl, cyclin D1, pml/rar alpha, AML1/MTG8,  
 CC E2A/P1 and ALL-1/AF-4. The composition is used for treating cancer.  
 CC The combination of antisense oligonucleotides has synergistically  
 CC enhanced ability to inhibit growth of cancer cells.  
 XX  
 SQ Sequence 3225 BP; 975 A; 687 C; 698 G; 865 T; 0 other;

Query Match 11.8%; Score 18; DB 19; Length 3225;  
 Best Local Similarity 100.0%; Pred. No. 9.3;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 97.ATTAAATTACTGAAATRAA 114  
 ||||||||||||||||  
 DB 2662 ATTAAATTACTGAAATRAA 2645

RESULT 13  
 ABL15492/C  
 ID ABL15492 standard; cDNA: 59967 BP.  
 XX  
 AC ABL15492;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 40958.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical; gene; ss.  
 OS  
 XX Drosophila melanogaster.  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PE 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR P-PSDB; ABB71389.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PT  
 XX  
 PS Claim 1; SEQ ID NO 40958; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176 ABL20511), expressed DNA  
 CC sequences (ABB7137-ABB72072).  
 CC (ABB7137-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 59967 BP; 17305 A; 12463 C; 12291 G; 17908 T; 0 other;

Query Match 11.8%; Score 18; DB 23; Length 59967;  
 Best Local Similarity 100.0%; Pred. No. 9.7;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 ACAGAGAACCCAGACCA 37  
 ||||||||||||||||  
 DB 21570 ACAGAGAACCCAGACCA 21553

RESULT 14  
 AA201319  
 ID AA201319 standard; DNA: 19 BP.  
 XX  
 AC AA201319;  
 XX  
 DT 27-SEP-1999 (first entry)  
 XX  
 DE PCR primer for PGI biallelic marker 99-140-130.  
 XX  
 DE PGI gene; biallelic marker; PCR primer; PGI-related biallelic marker;  
 KW cancer; prostate cancer; diagnosis; therapy; prostate specific antigen;  
 KW PSA; human; ss.  
 XX  
 OS Synthetic.  
 OS Homo sapiens.  
 XX  
 PN WO932644-A2.  
 XX  
 PD 01-JUL-1999.  
 XX  
 PE 22-DEC-1998; 98WO-IB02133.  
 XX  
 PR 09-SEP-1998; 98US-0099658.  
 PR 22-DEC-1997; 97US-0996306.  
 XX  
 PA (GSEST ) GENSET.  
 XX  
 PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen D;  
 XX  
 DR WPI; 1999-405178/34.  
 XX  
 PT Use of a prostate cancer associated gene and biallelic markers  
 PT derived from it  
 PT  
 XX  
 PS Claim 4; Page 368; 385pp; English.  
 XX  
 CC The invention relates to a mammalian PGI gene and protein, and a set of  
 CC PGI biallelic markers. The PGI polynucleotide and biallelic markers are  
 CC used in a hybridisation assay, a sequencing assay, or in an  
 CC allele-specific amplification assay for determining the identity of a  
 CC nucleotide at a PGI-related biallelic marker. The methods can be used to  
 CC detect and to assess the risk of developing cancer or prostate cancer.  
 CC Early-stage diagnosis of prostate cancer relies on prostate specific  
 CC antigen (PSA) dosage. However, the effectiveness of this is limited due  
 CC to its inability to discriminate between malignant and non-malignant  
 CC affections of the organ. A need exists for both a reliable diagnostic  
 CC procedure which would enable early-stage diagnosis, and for preventative  
 CC and curative treatments of the disease. The PGI gene can be used for  
 CC detection of prostate cancer, and the risk of developing it in the  
 CC future, and can also be used to determine therapies for the disease.  
 XX  
 SQ Sequence 19 BP; 9 A; 6 C; 3 G; 1 T; 0 other;

Query Match 11.1%; Score 17; DB 20; Length 19;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 AAGACGCTACAGACCA 52  
 ||||||||||||||||  
 DB 3 aagacgctacagacca 19

RESULT 15  
 AA201079  
 ID AA201079 standard; DNA: 47 BP.

XX	AA01079;
AC	
XX	27-SEP-1999 (first entry)
DT	
XX	
DE	Probe for human PGI biallelic marker 99-140-130.
XX	
KW	PGI gene; biallelic marker; PCR primer; PGI-related biallelic marker;
KW	cancer; prostate cancer; diagnosis; therapy; prostate specific antigen;
KW	PSA; human; ss.
XX	
OS	Synthetic.
OS	Homo sapiens.
XX	
PN	W09932644-A2.
XX	
PD	01-JUL-1999.
XX	
PF	22-DEC-1998; 98WO-IB02133.
XX	
PR	09-SEP-1998; 98US-0099658.
XX	
PR	22-DEC-1997; 97US-0996306.
XX	
PA	(GEST ) GENSET.
XX	
PI	Blumenfeld M, Bougueleret L, Chumakov I, Cohen D;
XX	
XX	WPI; 1999-405178/34.
DR	
XX	
PT	Use of a prostate cancer associated gene and biallelic markers
PT	derived from it
XX	
PS	Claim 4; Page 313; 385pp; English.
XX	
CC	The invention relates to a mammalian PGI gene and protein, and a set of
CC	PGI biallelic markers. The PGI polynucleotide and biallelic markers are
CC	used in a hybridisation assay, a sequencing assay, or in an
CC	allele-specific amplification assay for determining the identity of a
CC	nucleotide at a PGI-related biallelic marker. The methods can be used to
CC	detect and to assess the risk of developing cancer or prostate cancer.
CC	Early-stage diagnosis of prostate cancer relies on prostate specific
CC	antigen (PSA) dosage. However, the effectiveness of this is limited due
CC	to its inability to discriminate between malignant and non-malignant
CC	affections of the organ. A need exists for both a reliable diagnostic
CC	procedure which would enable early-stage diagnosis, and for preventative
CC	and curative treatments of the disease. The PGI gene can be used for
CC	detection of prostate cancer, and the risk of developing it in the
CC	future, and can also be used to determine therapies for the disease.
XX	
XX	Sequence 47 BP; 17 A; 10 C; 10 G; 10 T; 0 other;

Query Match	11.1%	Score 17	DB 20	Length 47
Best Local Similarity	100.0%	Pred. No. 29		
Matches	17	Conservative	0	Mismatches 0; Indels 0; Gaps 0
QY	36	AAAGCAGCTACAGACCA	52	
DB	7	aaagcagctacagacca	23	

Search completed: August 3, 2002, 01:50:07  
Job time: 32192 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:33 ; Search time 7016.61 Seconds

(without alignments)  
294.307 Million cell updates/sec

Title: US-08-973-363-3

Sequence: 1 ATTTCACCTGATGATCCAGA.....CACAAAGCCTGCTGTCGCA 153

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size: 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em.estbta:\*  
2: em.estbthum:\*  
3: em.estbnu:\*  
4: em.estbmu:\*  
5: em.estbvt:\*  
6: em.estbpl:\*  
7: em.estbro:\*  
8: em.estbrc:\*  
9: gb.estb1:\*  
10: gb.estb2:\*  
11: gb.hic:\*  
12: gb.gss:\*  
13: em.gss.hum:\*  
14: em.gss.liv:\*  
15: em.gss.pln:\*  
16: em.gss.vtl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	16.3	619	9	BB155356
2	25	16.3	619	9	BB155356
3	23	15.0	430	9	BB461065
4	23	15.0	430	9	AL601246
5	23	15.0	438	9	BB830730
6	23	15.0	446	9	BB834922
7	23	15.0	547	9	AI890775
8	23	15.0	686	9	AM997058
9	23	15.0	821	10	BF239667
10	23	15.0	866	9	AUI25712
11	21	13.7	1028	10	BE895133
12	20	13.1	693	12	A2570983
13	20	13.1	693	12	BB351821
14	19	12.4	505	12	AG128961
15	19	12.4	527	12	AQ453778
16	19	12.4	558	12	A2486207
17	19	12.4	573	12	A2392020
					AQ988250
					RPCT-23-3

18	19	12.4	695	12	A2963945
19	19	12.4	895	10	BE971206
20	19	12.4	944	12	A2200903
21	18	11.8	110	10	221483
22	18	11.8	157	12	BB4318
23	18	11.8	167	10	BE842996
24	18	11.8	225	9	AV317344
25	18	11.8	225	9	AV317684
26	18	11.8	233	9	AV312872
27	18	11.8	233	9	AV375787
28	18	11.8	235	9	BB164147
29	18	11.8	276	9	AV222464
30	18	11.8	276	10	BE769884
31	18	11.8	291	9	BB030255
32	18	11.8	293	9	BB031183
33	18	11.8	301	9	BB402744
34	18	11.8	302	9	BB560958
35	18	11.8	303	9	BB17675
36	18	11.8	314	9	AA354813
37	18	11.8	322	10	BF289372
38	18	11.8	323	10	BB137014
39	18	11.8	323	9	AI902726
40	18	11.8	416	9	BB810785
41	18	11.8	424	12	AO816352
42	18	11.8	426	9	AA267899
43	18	11.8	437	9	AA955076
44	18	11.8	438	9	AA234796
45	18	11.8	438	9	AA234796

## ALIGNMENTS

RESULT 1  
LOCUS BB155356  
DEFINITION BB155356 RIKEN full-length cDNA clone A13002416 3' similar to I10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION BB155356.2 GI:16268254  
VERSION BB155356  
KEYWORDS EST  
SOURCE house mouse,  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. (bases 1 to 619)

On Jun 29, 2000 this sequence version replaced gi:881186.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp,  
URL: http://genome-gsc.riken.go.jp/  
Carinchi, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagui, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, R.,  
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura

S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamanake, I., Atzawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

BASE COUNT	194 a	118 c	161 g	146 t
ORIGIN				

Query Match	16.3%;	Score 25;	DB 9;	Length 619;
Best Local Similarity	100.0%;	Pred. No. 0.0061;		
Matches 25;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	74	TACAGACCCGTCAGACTACCTCAT	98
Db	532	TACAGACCCGTCAGACTACCTCAT	556

Db 532 TACAGACCCGTCGACACTACCTCAT 556

RESULT	2
LOCUS	BB461065
DEFINITION	BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion Mus musculus cDNA clone D13007OB13 3' similar to LI0410 mouse DNA-binding protein (CND-1) mRNA, mRNA sequence.
ACCESSION	BB461065
VERSION	BB461065.2 GI:16426612
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE  
AUTHORS

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus. 1 (bases 1 to 660)
AUTHORS	Arakawa, T., Carrinno, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, H., Ozakaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinnagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takehashi, F., Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.
TITLE	RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
JOURNAL	Unpublished (2001)
COMMENT	On Jul 21, 2000 this sequence version replaced g1:9356558.

**FEATURES**  
**source**



RESULT 5  
BB834922 446 bp mRNA linear EST 19-NOV-2001  
LOCUS BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
DEFINITION Jy9-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA  
sequence.  
ACCESSION BB834922  
VERSION BB834922.1 GI:17013165  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 446)  
AUTHORS Akimura,T., Arikawa,T., Carninci,P., Furuno,M., Hanagaki,T.,  
Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,  
Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,  
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,  
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,  
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,  
A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.,  
Watanhki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
2001)  
TITLE Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
URL: http://genome.gsc.riken.go.jp/  
Email: genome-res@sc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,  
M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagl,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watanhki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.  
FEATURES  
Location/Qualifiers  
source 1..446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930033J21"  
/clone.lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jy9-MC(B)"  
BASE COUNT 139 a 99 c 108 g 100 t  
ORIGIN  
Query Match 15.0%; Score 23; DB 9; Length 446;  
Best Local Similarity 100.0%; Pred. No. 0.072;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
74 TACAGACCGTGCAGACTACCTC 96  
|||||

Db 352 TACAGACCGTGCAGACTACCTC 374  
RESULT 6  
A1890775/c 547 bp mRNA linear EST 07-MAR-2000  
LOCUS A1890775 NCI-CGAP.Ut2 Homo sapiens cDNA clone IMAGE:2443725 3',  
DEFINITION similar to SW:CHD1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1; , mRNA sequence.  
ACCESSION A1890775  
VERSION A1890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 547)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/INL at:  
www.bio.lnl.gov/bdrp/image/image.html  
Insert Length: 1924 Std Error: 0.00  
Seq primer: -40UP from G1bpco  
High quality sequence stop: 418.  
FEATURES  
Location/Qualifiers  
source 1..547  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2443725"  
/clone.lib="NCI-CGAP Ut2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; site\_1: SalI;  
site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 kb. Life Technologies catalog #: 11539-012"  
BASE COUNT 114 a 118 c 85 g 230 t  
ORIGIN  
Query Match 15.0%; Score 23; DB 9; Length 547;  
Best Local Similarity 100.0%; Pred. No. 0.074;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
76 CAGACCCGTGCAGACTACCTCAT 98  
|||||  
Db 333 CAGACCCGTGCAGACTACCTCAT 311  
RESULT 7  
AW997058 686 bp mRNA linear EST 05-JUN-2000  
LOCUS AW997058 OV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION AW997058  
VERSION AW997058.1 GI:8257292  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 686)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,



Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bais,G.S., Simpson,D.H.,  
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 'M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 COMMENT  
 JOURNAL  
 MEDLINE  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the PAPBS/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=kt2-qv3-BN0047-150  
 400-152-c03&f3=2000-04-15&f4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 19  
 High quality sequence stop: 678.  
 Location/Qualifiers  
 1..686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="BN0047"  
 /dev\_stage="Adult"  
 /note="Organ: breast\_normal; Vector: puc18; Site:1: Sma1;  
 Site:2: Sma1; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 BASE COUNT 165 a 154 c 126 g 241 t  
 ORIGIN  
 Query Match 15.0%; Score 23; DB 9; Length 686;  
 Best Local Similarity 100.0%; Pred. No. 0.077; 0; Indels 0; Gaps 0;  
 Matches 23; Conservative 0; Mismatches 0;  
 Oy 76 CAGACCCGTCGACGACTCCTCAT 98  
 ||||||||||||||||||||  
 Db 83 CAGACCCGTCGACGACTCCTCAT 61  
 RESULT 8  
 BF239967 821 bp mRNA linear EST 14-NOV-2000  
 LOCUS 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BF239967  
 VERSION BF239967.1 GI:11153890  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 821)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs@mail.nih.gov  
 Tissue Procurement: ATCC  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

Found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: LCM1033 row: k column: 18  
 High quality sequence stop: 562.  
 Location/Qualifiers  
 1..821  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="IMAGE:4133129"  
 /clone\_lib="NIH\_MGC\_54"  
 /issue\_type="from Chronic myelogenous leukemia"  
 /lab\_host="pMD108 (T1 phage-resistant)"  
 /note="Organ: bone marrow; Vector: pMDR-LIB (Clontech);  
 Site:1: SfiI (ggcgccggcgcc); Site:2: SfiI (ggcgatggcc  
 ); Double-stranded cDNA was prepared from cell line RNA.  
 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CAGGCGCATTTTGACC-3' and 3' adaptor  
 sequence: 5'-ATTCTAGAGGCGGCGGCGGCGCATG-dt(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."  
 BASE COUNT 345 a 149 c 190 g 137 t  
 ORIGIN  
 Query Match 15.0%; Score 23; DB 10; Length 821;  
 Best Local Similarity 100.0%; Pred. No. 0.079; 0; Indels 0; Gaps 0;  
 Matches 23; Conservative 0; Mismatches 0;  
 Oy 76 CAGACCCGTCGACGACTCCTCAT 98  
 ||||||||||||||||||||  
 Db 33 CAGACCCGTCGACGACTCCTCAT 55  
 RESULT 9  
 AUI25712 866 bp mRNA linear EST 23-OCT-2000  
 LOCUS AUI25712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
 DEFINITION Sequence.  
 ACCESSION AUI25712  
 VERSION AUI25712.1 GI:10950428  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 866)  
 Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
 Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
 Isogai,T.  
 HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
 Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
 'Y., Sugano,S., Isogai,T.)  
 Unpublished (2000)  
 COMMENT Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3951  
 Fax: 81-438-52-3952  
 Email: genomics@hri.co.jp  
 HRI human cDNA project: 5'- 5' and one pass sequencing: Helix  
 Research Institute, cDNA library construction: Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 Location/Qualifiers  
 1..866  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="NT2RM4002061"  
 /clone\_lib="NT2RM4"

```

/cell_type="teratocarcinoma"
/cell_line="NT2"
/Note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"
BASE COUNT      312 a      149 c      196 g      207 t      2 others
ORIGIN

Query Match      15.0%; Score 23; DB 9; Length 866;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 CAGACCCGTCGACACTACCTCAT 98
|||||
Db 495 CAGACCCGTCGACACTACCTCAT 517

RESULT 10
LOCUS      BE895133      1028 bp      mRNA      linear      EST 20-OCT-2000
DEFINITION 601436060F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3921087 5',
ACCESSION  BE895133
VERSION     BE895133.1
KEYWORDS   GI:10358221
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 1028)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cygabs-remail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM9753 row: h column: 16
            High quality sequence stop: 488.
            Location/Qualifiers
                1..1028
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3921087"
                /clone_1ib="NIH_MGC_72"
                /tissue_type="melanotic melanoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: skin; Vector: pCMV-Sport6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 2 kb. Library constructed by Life
                Technologies."
BASE COUNT      387 a      205 c      238 g      198 t
ORIGIN

Query Match      15.0%; Score 23; DB 10; Length 1028;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 76 CAGACCCGTCGACACTACCTCAT 98
|||||
Db 176 CAGACCCGTCGACACTACCTCAT 198

RESULT 11
LOCUS      A2570983      664 bp      DNA      linear      GSS 15-MAY-2001
DEFINITION 281FV04 PV MBN #30 Plasmodium vivax genomic 3', DNA sequence.

```

```

ACCESSION  A2570983
VERSION     A2570983.1
KEYWORDS   GI:13982617
SOURCE     GSS.
            malaria parasite P. vivax.
ORGANISM   Plasmodium vivax
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE  1 (bases 1 to 664)
            Carlton J.M.-R. and Dame J.B.
            The Plasmodium vivax and P. berghei gene sequence tag projects
            Parasitol. Today 16 (10), 409 (2000)
JOURNAL    Contact: Dame JB
            Dept. of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.ufl.edu
            Seq primer: M13(-20) forward
            Class: shotgun.
            Location/Qualifiers
                1..664
                /organism="Plasmodium vivax"
                /strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
                497-598)"
                /db_xref="taxon:5855"
                /clone_1ib="PV MBN #30"
                /dev_stage="asexual blood forms"
                /lab_host="Saimiri boliviensis"
                /note="Vector: pBluescript SK(+) vector DNA, phagemid
                excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V;
                Host leukocytes were extracted from P. vivax infected
                blood using the following methods: first, infected blood
                was activated by the addition of 0.5 ml of ADP (40mg/ml)
                per 10 ml blood. Then blood was passed over a column of
                acid washed 0.1 mm glass beads, then through a Plasmidpur
                filter, followed by passage through a column of pre-wet
                Whatman CF11 powder (1:2 ratio volume of blood to CF11),
                and finally centrifuged through a 50% Percoll density
                cushion. Purified DNA was digested with mung bean nuclease
                in the presence of 44% formamide at 500C as described
                (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988.
                Nucleic Acids Research 16:6883-6896). Digested DNA was
                blunt-ended using T4 DNA Polymerase and size fractionated
                over a Sepharose CL-2B column. Fractions in the size range
                500bp-4kb were ligated into the Eco RV site of pBluescript
                SK(+), and E. coli XL-10 Gold transformed with the
                ligation mixture."
BASE COUNT      197 a      159 c      115 g      193 t
ORIGIN

Query Match      13.7%; Score 21; DB 12; Length 664;
Best Local Similarity 100.0%; Pred. No. 0.93;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 GAAAGGAGCAACAAGGCTTG 145
|||||
Db 402 GAAAGGAGCAACAAGGCTTG 422

RESULT 12
LOCUS      BH351821      693 bp      DNA      linear      GSS 03-DEC-2001
DEFINITION CH230-160016.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
ACCESSION  BH351821
VERSION     BH351821
KEYWORDS   GI:17282555
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.

```

REFERENCE 1 (bases 1 to 693)  
 AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C. M.  
 TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_GSSs: CH230-160016.TJ  
 Contact: Shaying Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability please contact Pieter de Jong (pdejong@tigr.org).  
 Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/orering\_information.htm). BAC end page: http://www.tigr.org/cdb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 160  
 Row: 0  
 Column: 16  
 Seq primer: T7

FEATURES  
 source  
 1..693  
 /organism="Rattus norvegicus"  
 /strain="BN/SSNhsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone\_id="CH230-160016"  
 /clone\_1lb="CHORI-230 Segment 1"  
 /cell\_type="Brain"  
 /note="Vector: PTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI; CHORI-230 Rat (BN/SSNhsd/MCW) BAC library produced by Pieter de Jong"

BASE COUNT 285 a 147 c 158 g 103 t

ORIGIN

Query Match 13.1%; Score 20; DB 12; Length 693;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 CAGGAGGAGGACGACGAGG 141  
 |||||||||||||||||||  
 Db 597 CAGGAGGAGGACGACGAGG 616

RESULT 13  
 AG128961 838 bp DNA linear GSS 04-NOV-2001  
 LOCUS Pan troglodytes DNA, clone: PTB-140E13.F, genomic survey sequence.  
 DEFINITION AG128961  
 ACCESSION AG128961.1 GI:16658126  
 VERSION GSS: GSS (genome survey sequence).  
 KEYWORDS Pan troglodytes male lymphoblast DNA, clone\_1lb:PTB Chimpanzee Male BAC library clone:PTB-140E13.F.  
 SOURCE Pan troglodytes  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.  
 REFERENCE 1 (sites)  
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE BAC end sequences of library PTB  
 JOURNAL Unpublished  
 COMMENT 2 (bases 1 to 838)  
 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.  
 Direct Submission  
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:chimpesgsc.riken.go.jp, url:http://hnp.gsc.riken.go.jp/,

COMMENT Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the Rad process and may have higher chance of clone tracking errors.  
 PRIMERS  
 Sequencing: -21M13  
 LIBRARY  
 Vector : pRS145  
 R.Site 1 : SacI  
 R.Site 2 : SacI.  
 location/Qualifiers  
 source  
 1..838  
 /organism="Pan troglodytes"  
 /db\_xref="taxon:9538"  
 /clone\_id="PTB-140E13.F"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone\_1lb="PTB Chimpanzee Male BAC library"

BASE COUNT 286 a 276 c 177 g 44 t 55 others

ORIGIN

Query Match 13.1%; Score 20; DB 12; Length 838;  
 Best Local Similarity 100.0%; Pred. No. 3.4;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GAACCCGAGGCAAGCAGC 43  
 |||||||||||||||||||  
 Db 458 GAACCCGAGGCAAGCAGC 477

RESULT 14  
 A0453778 505 bp DNA linear GSS 21-APR-1999  
 LOCUS HS-5170.AL.D02.T7A RPCI-11 Human Male BAC Library Homo sapiens  
 DEFINITION A0453778  
 ACCESSION A0453778.1 GI:4594948  
 VERSION GSS:  
 KEYWORDS human.  
 SOURCE Homo sapiens  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 505)  
 AUTHORS Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M. D. and Hood, L.  
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 COMMENT Contact: Mahairas G.G., Wallace J.C., Hood L.  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pdejong@med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu  
 Plate: 746  
 Row: G  
 Column: 3  
 Seq primer: T7  
 Class: BAC ends  
 location/Qualifiers  
 source  
 1..505  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="Plate-746 Col-3 Row-G"

```

/c/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/Note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT      197 a      77 c      107 g      119 t      5 others
ORIGIN

Query Match      12.4%; Score 19; DB 12; Length 505;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 123 AAGAAGGAGCACAAGG 141
Db 132 AAGAAGGAGCACAAGG 150

RESULT 15
AZ486207/c 527 bp DNA linear GSS 05-OCT-2000
LOCUS 1M0314M02F Mouse 10kb plasmid UGCC1M library Mus musculus genomic
DEFINITION
Clone UGCC1M0314M02 F, DNA sequence.
ACCESSION
AZ486207
VERSION
AZ486207.1 GI:10652755
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
1 (bases 1 to 527)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10Kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., StC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0314 Row: M Column: 02
Seq primer: CGTGTGTAAGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 527.
Location/Qualifiers
1..527
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCC1M0314M02"
/c/clone_lib="Mouse 10Kb plasmid UGCC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/Note="Vector: PMD42uv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel

```

```

electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      174 a      88 c      87 g      178 t
ORIGIN

Query Match      12.4%; Score 19; DB 12; Length 527;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 130 GAAGCACAAGGCTGCTG 148
Db 404 GAAGCACAAGGCTGCTG 386

```


Search completed: August 2, 2002, 22:41:37  
Job time: 30152 sec

Mon Aug 5 11:51:50 2002

us-08-973-363-3.011.rst

---

Page 9







Query Match 100.0%; Score 153; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 1.2e-66;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTTACTGATGACCCAGATAAGAAACACAGCAAGAGTTGGACACCAAGAAACCA 60  
|||||  
1 ATTTACTGATGACCCAGATAAGAAACACAGCAAGAGTTGGACACCAAGAAACCA 60

QY 61 CAGGCAAGAGAGTTGCAGAGACCCCTGCAGATTACTCTATTAAATTACTGATTAAGACCTT 120  
|||||  
61 CAGGCAAGAGAGTTGCAGAGACCCCTGCAGATTACTCTATTAAATTACTGATTAAGACCTT 120

Db 121 GCAGAAAGAGAGTTGCAGAGACCTTACTGATGCA 153  
|||||  
121 GCAGAAAGAGAGTTGCAGAGACCTTACTGATGCA 153

RESULT 2  
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 4 from Patent WO9639505.  
ACCESSION A58685  
VERSION A58685.1 GI:3714248  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source 1. .153  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 56 a 36 c 31 g 30 t  
ORIGIN

Query Match 34.6%; Score 53; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 3.9e-16;  
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 CAGACCCGTGCAGATTACTCTATTAAATTACTGATTAAGACCTTGCAGAGAA 128  
|||||  
76 CAGACCCGTGCAGATTACTCTATTAAATTACTGATTAAGACCTTGCAGAGAA 128

Db 76 CAGACCCGTGCAGATTACTCTATTAAATTACTGATTAAGACCTTGCAGAGAA 128

RESULT 3  
LOCUS A58696 1316 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 15 from Patent WO9639505.  
ACCESSION A58696  
VERSION A58696.1 GI:3714253  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1316)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 15 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source 1. .1316  
/organism="unidentified"

BASE COUNT 493 a 205 c 308 g 304 t 6 others  
ORIGIN /db\_xref="taxon:32644"

Query Match 30.7%; Score 47; DB 6; Length 1316;  
Best Local Similarity 100.0%; Pred. No. 2.9e-13;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 CGTGCAGATTACTCTATTAAATTACTGATTAAGACCTTGCAGAGAA 128  
|||||  
1189 CGTGCAGATTACTCTATTAAATTACTGATTAAGACCTTGCAGAGAA 1235

Db 1189 CGTGCAGATTACTCTATTAAATTACTGATTAAGACCTTGCAGAGAA 1235

RESULT 4  
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 3 from Patent WO9639505.  
ACCESSION A58684  
VERSION A58684.1 GI:3714247  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source 1. .153  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 58 a 40 c 31 g 24 t  
ORIGIN

Query Match 24.8%; Score 38; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 1.5e-08;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACCTCATTAATTACTGATTAAGACCTTGCAGAGAA 128  
|||||  
91 TACCTCATTAATTACTGATTAAGACCTTGCAGAGAA 128

Db 91 TACCTCATTAATTACTGATTAAGACCTTGCAGAGAA 128

RESULT 5  
LOCUS D14316 2292 bp mRNA linear VRT 03-FEB-1999  
DEFINITION delta-crystallin enhancer binding protein, complete sequence.  
ACCESSION D14316  
VERSION D14316.1 GI:391639  
KEYWORDS ORF2.  
SOURCE Gallus gallus (library: lambda gt11) 13 day embryo lens cDNA to mRNA, clone vfl1.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 2292)  
AUTHORS Funahashi,J.  
TITLE Direct Submission  
JOURNAL Submitted (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel. Aging and Cancer, Tohoku Univ.; 4-1 Seiry-machi, Aoba-ku, Sendai 980-77,  
Japan (Tel:022-272-9499, Fax:022-272-3982)  
2 (sites)  
REFERENCE Funahashi,J., Sekido,R., Murai,K., Kamachi,Y. and Kondoh,H.  
AUTHORS Delta-crystallin enhancer binding protein delta EFL is a zinc  
TITLE finger-homeodomain protein implicated in postgastrulation  
JOURNAL embryogenesis  
Development 119 (2), 433-446 (1993)



MEDLINE 94116444  
REFERENCE 3 (bases 1 to 2292)  
AUTHORS Funahashi,J.  
JOURNAL Unpublished (1994)  
FEATURES  
source Location/Qualifiers  
1..2292  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/clone\_lib="lambda g11"  
/dev\_stage="13 day embryo"  
257..1939  
/note="beta-gal fusion protein binds to blocks 10 to 3 of  
HN fragment of delta-crystallin enhancer."  
/codon\_start=1  
/product="ORF2"  
/protein\_id="BAA03262.1"  
/db\_xref="GI:391640"  
/translation="MELKKCCNHCYLIRPDNDNEFYKNOALQHLIRSSGKLILLDL  
LIRLRGNRVLIFSOWVRMLDILAYLRKRFPPORLDSIKGELRQALDHRNAG  
SEDCPLSTFRAGGGINLASADVYIPDSNDPNDLQOARAHICQKQVAVYL  
VTGKSVLEOLLEPARKKVDLHVLRMTGKLYLHTGSSSTPPKKELSALIK  
FGAELEKPEGEPOEDMDIDELIKRAETENPGLIPGELLSQKANSND  
PDDELEPNSRNMEETIPESORRIIEEPOKLEIYMLPRMNCALQISNGE  
GRSRSRSGSDSITERKRNKGRRTIRFNTKGSDEIRNFKIKRGGP  
LERIDAVARAEIVKSETDLRLGELVNGCTRLKNSGGERAGGLGAVGPTE  
KISGVNNAVLITSEELAPLHKSIPSPERKRTYIPCHTAHFIDIMKEDSDN  
LIVEITAYGSKQCKSRSSQINTEILIPDPRTPRONSYPVOTISLNT"

BASE COUNT 827 a 385 c 546 g 534 t  
ORIGIN

Query Match 24.8%; Score 38; DB 5; Length 2292;  
Best Local Similarity 100.0%; Pred. No. 9.2e-09;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACCTCATTAATCTGATTAAGACCTTGCAAGAAA 128  
|||||  
Db 1921 TACCTCATTAATCTGATTAAGACCTTGCAAGAAA 1958  
|||||

RESULT 6  
LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 10 from Patent W096939505.  
ACCESSION A58691  
VERSION A58691.1 GI:3714250  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 6608)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE Avian GHD Genes and Their Use in Methods for Sex Identification in  
JOURNAL Birds  
PATENT: WO 9639505-A 10 12-DEC-1996;  
ISIS INNOVATION (GB)  
COMMENT Other Publication AU 5906966 961224.  
FEATURES  
source Location/Qualifiers  
1..6608  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN

Query Match 24.8%; Score 38; DB 6; Length 6608;  
Best Local Similarity 100.0%; Pred. No. 7.7e-09;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 TACCTCATTAATCTGATTAAGACCTTGCAAGAAA 128  
|||||  
Db 4140 TACCTCATTAATCTGATTAAGACCTTGCAAGAAA 4177  
|||||

RESULT 7  
AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
LOCUS AF004397  
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the Z chromosome  
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete  
cds.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE  
ORGANISM chicken.  
Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Phasianidae; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 6872)  
A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus  
cds.  
REFERENCE 1 (bases 1 to 6872)  
AUTHORS Griffiths,R. and Korn,R.M.  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAR-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK  
FEATURES  
source Location/Qualifiers  
1..6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/gene="CHD-Z"  
/db\_xref="GI:2501845"  
/translation="MNGHSDSESVNNSGSSGSSRDSDSGSASGSSGSSGSSGSS  
SOGSSDSESGSSESDTSREK KOVAKPPKADSEFPKSPSILVORSAV  
LKQOQOQKAASDSDGSESDSSADSSSETRKKKHDEDMOQSGSVGTS  
DSEASDSDGSCSESESDYEPKAKVSRKPPERTKPRGSKSGKQRLDSSEEE  
DDDEDYDKGSRQATVAVSYKAEKERTSDDLTEVGGEPVOTEDDEFTKRFND  
SRIGKRGATGASTTIYAVADGDNAGEKSEKLEIYLIKRRKQSHLHNWFEET  
LKQWVKNMKNLKNYKKNQPEFKWLNKASPEVEYVCOQELDKQVYERTI  
AHSNKSAGCPDYCKRQGLPYSGSMEDGALIAKROAIDEPYRSQSTTFED  
CVYIKQRPYALKQPSYSGHSEIEIPYOLINGMLASRCKSGCITLDEMGIC  
KITQITISFLAVPHHQILGPELRLVPLSTLWQREIQYAPQNNVYVLDITSRN  
MRTHEMHHPQIKLKFNTLITTEILDKSLGSLGMATIGVDEHAKRDESLAS  
RLIDPKSNRHLITGTPLONSLKLNSLDLTPKTSMEDEEERKGEVITAS  
LRLKEPLRLRYKRCVKSLEPAVBOILMMESALQVYKATLIRNKALKSGSKG  
STGSLNIMMELKCCNHCYLIRPDNDNEFYKNOALQHLIRSSGKLILLDLIRK  
EGRNVLIFSOWVRMLDILAYLRKRFPPORLDSIKGELRQALDHRNAGSDFC  
FLLSTRAGGLGINLASADVYIPDSNDPNDLQOARAHICQKQVAVYLVKYS  
VEEDILERAKKVDLHVLRMTGKLYLHTGSSSTPPKKELSALIKFGAE  
LEDEGEEOEDMDIDELIKRAETENPGLIPGELLSQKANSNDPDDELE  
LEPERNSRNMEETIPESORRIIEEPOKLEIYMLPRMNCALQISNGEGRSR  
SRYSGSDSITERKRNKGRRTIRFNTKGSDEIRNFKIKRGGP  
AVARAEIVKSETDLRLGELVNGCTRLKNSGGERAGGLGAVGPTEKIS  
QVNAVLITSEELAPLHKSIPSPERKRTYIPCHTAHFIDIMKEDSDNLTGVI  
VEYGSGWEMTKMDPLSTQKILIPDDPKRSPQAKOQJTRADYILILKNDLARK  
FLAGAGSNKRRKTRNKNMKKASKIKEIKSDSSQPEKDEDEEDNKLEYSK  
HLAKKIKTEKNEKEPEPDIGIKKEAEKRETERKREKREKREKREKLEKED  
NKEKREKVEKSTOKEVKREKVEKSEKSKITPLDTPVHITLITKSEKPEIS  
ESELHOKTEFVCKEEMRVKALKOLDRPEKLESEKRLKFOGCLIKIGHITELC  
KEYTNPDIKQMRKNIMLIVKFTFEPDARLHKYKAIKKRQSGOHNDOINISNV  
TVIINPDEVERLEKETTNNHDSRSOSYSDRNLQYHDKHKKRQCGAVKSSRRPY  
SAFSGKRRHMDHYKQDSRRYSKSKRKRLDSOYHDKHKKRQCGAVKSSRRPY  
SDHRISDHSRSTSEYSHKSSRDYSHSDMOMHRRASSGSRPSRLDQSRPYGSHRSH  
HSPSHSHSHKSTPETHWSSRKT"

```

misc-feature      4341..4604
                  /gene="CHD-2"
                  /note="short insert found in longer variant mRNA of CHD-2"
BASE COUNT      2446 a 1223 c 1520 g 1683 t
ORIGIN

Query Match      24.8%; Score 38; DB 5; Length 6872;
Best Local Similarity 100.0%; Pred. No. 7.7e-09;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 91 TACCTCATTAATTAAGACCTTGCAAGAA 128
|||||
Db 4140 TACCTCATTAATTAAGACCTTGCAAGAA 4177

RESULT 8
LOCUS      MM091538      959 bp mRNA linear ROD 12-APR-1997
DEFINITION Mus musculus vesicle trafficking protein sec22b mRNA, complete cds.
ACCESSION  U91538
VERSION     U91538.1 GI:1907385
KEYWORDS
SOURCE      house mouse.
ORGANISM    Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 959)
AUTHORS    Hay, J.C., Chao, D.S., Kuo, C.S. and Scheller, R.H.
TITLE      Protein interactions regulating vesicle transport between the
            endoplasmic reticulum and Golgi apparatus in mammalian cells
            Cell 89 (1), 149-158 (1997)
JOURNAL    97248495
MEDLINE    2 (bases 1 to 959)
REFERENCE   Hay, J.C., Chao, D.S., Kuo, C.S. and Scheller, R.H.
AUTHORS    Direct Submission
TITLE      Submitted (27-FEB-1997) Molecular and Cellular Physiology, Stanford
            University, B155 Beckman Center, Stanford, CA 94305, USA
FEATURES
source      1..959
            /organism="Mus musculus"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /tissue_type="Placenta"
            77..724
            /note="similar to Sec22"
            /codon_start=1
            /product="vesicle trafficking protein sec22b"
            /protein_id="AAC53130.1"
            /db_xref="gi:1907386"
            /translation="MVLMTIARVADGLPLASMOEDSGRDLQOYOSQAKLFRKL
            NEOSPRTCTLGAMTFHYITIEOGVCYLVCEAAPPKLAFAVLEDLSEPEQHGK
            VPTVSRPYSFIETDFDIOKTKRLYIDSRARRLGSINTELQVQIRVIANIEVYORG
            EALSRLDSKANLISLSKRYODATYLMKRSYAKLAVAVFIMLYVVRPWL"

BASE COUNT   241 a 219 c 261 g 238 t
ORIGIN

Query Match      13.7%; Score 21; DB 10; Length 959;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131
|||||
Db 835 TAAAGACCTTGCAAGAAAGA 855

RESULT 9
LOCUS      BC009024      1339 bp mRNA linear ROD 12-JUL-2001
DEFINITION Mus musculus; SEC22, vesicle trafficking protein (S.
            cerevisiae)-like 1, clone MGC:5885 IMAGE:3481731, mRNA, complete
            cds.

```

```

ACCESSION      BC009024
VERSION        BC009024.1 GI:14290511
KEYWORDS
SOURCE         house mouse.
ORGANISM       Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 1339)
AUTHORS        Strausberg, R.
TITLE          Direct Submission
JOURNAL        Submitted (30-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK         NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT         Contact: MGC help desk
            Email: cgephs-remail.nih.gov
            Tissue Procurement: Lotnar Hennighausen Ph.D., Robin Humphreys
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLJ)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: villalona@bcm.tmc.edu.
            Villalona, D.K., Luna, R.A., Hale, S.M., Huilyk, S., Lu, X., Garcia,
            A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
            Muzny, D.M., Gibbs, R.A.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/BLN: at: http://Image.llnl.gov
            Series: IRAP Plate: 6 Row: k Column: 24
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 1907385.
FEATURES
source      1..1339
            /organism="Mus musculus"
            /db_xref="LocusID:9554"
            /db_xref="taxon:10090"
            /clone="MGC:5885 IMAGE:3481731"
            /tissue_type="Mammary tumor. MAP-1Gf alpha model. 7 months
            old, gross tissue."
            /clone_11b="NCI-CGAP_Mam5"
            /lab_host="DH10B"
            /note="Vector: PCMV-SPORT6"
            82..729
            /codon_start=1
            /product="SEC22, vesicle trafficking protein (S.
            cerevisiae)-like 1"
            /protein_id="AAH09024.1"
            /db_xref="gi:14290512"
            /translation="MVLMTIARVADGLPLASMOEDSGRDLQOYOSQAKLFRKL
            NEOSPRTCTLGAMTFHYITIEOGVCYLVCEAAPPKLAFAVLEDLSEPEQHGK
            VPTVSRPYSFIETDFDIOKTKRLYIDSRARRLGSINTELQVQIRVIANIEVYORG
            EALSRLDSKANLISLSKRYODATYLMKRSYAKLAVAVFIMLYVVRPWL"

BASE COUNT   347 a 291 c 344 g 357 t
ORIGIN

Query Match      13.7%; Score 21; DB 10; Length 1339;
Best Local Similarity 100.0%; Pred. No. 3.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131
|||||
Db 840 TAAAGACCTTGCAAGAAAGA 860

RESULT 10
LOCUS      BC001364      1491 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, SEC22, vesicle trafficking protein (S.
            cerevisiae)-like 1, clone MGC:1960 IMAGE:3051087, mRNA, complete
            cds.

```

ACCESSION BC001364  
VERSION 1  
KEYWORDS MGC  
SOURCE Human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 1491)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DMP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Beaudet, B., Blakeley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, O.L., Masello, C., Mastrian, S.D., McCluskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantrifop, S., Thomas, P.J., Thompson, E.E., Touchman, J.W., Tsurgou, C., Vogt, U.L., Walker, M.A., Zhang, L.-H. and Green, E.D.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL: <http://image.llnl.gov>  
Series: IRAL plate 4 Row: C Column: 13  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 335139.  
FEATURES  
source  
1..1491  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="MGC:1960 IMAGE:3051087"  
/tissue\_type="Skin, melanotic melanoma."  
/clone\_1b="NIR\_MGC\_20"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
129..776  
/codon\_start=1  
/product="SEC22, vesicle trafficking protein (S. cerevisiae)-like 1"  
/protein\_id="AAH01364.1"  
/db\_xref="GI:12655033"  
/translation="MVLITMTARVADGLPLASMOEDQSGRDLDQYQSOAKLFRKL  
NEOSTPCTLEAGMTFYIIEGQVCYLVCAGAPKKLAFAVLELSEFEDHGK  
VPTVSRYSFIEEDTETIKTKLIDSRARILGSIINLEQDQVRLIMVINEVYLGK  
FALALDSKNNISLSIKRYRODAKYLNRRSTYKLAIAVAVFIMLYVRFMWL"  
BASE COUNT 428 a 289 c 342 g 432 t  
ORIGIN  
Query Match 13.7%; Score 21; DB 9; Length 1491;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AL359758  
LOCUS AL359758 85836 bp DNA linear PRI 13-SEP-2001  
DEFINITION Human DNA sequence from clone RP11-544024 on chromosome 1, complete  
sequence.  
ACCESSION AL359758  
VERSION AL359758.14 GI:15620621  
KEYWORDS HTG.  
SOURCE Human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 85836)  
AUTHORS Lovell, J.  
TITLE Direct Submission  
JOURNAL Submitted (13-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
Requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Sep 14, 2001 this sequence version replaced gi:13897376.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL, S., SWISSPROT, Tr., TrEMBL, Wp., WormPeP, information on the WormPeP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch1>  
RP11-544024 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/dacpac/home.htm>  
VECTOR: pBAC3.6  
IMPORTANT: This sequence is not the entire insert of clone RP11-544024 it may be shorter because we sequence overlapping sections only once, except for a short overlap.  
The true left end of clone RP11-544024 is at 1 in this sequence. The true left end of clone RP4-646P11 is at 85737 in this sequence.  
FEATURES  
source  
1..85836  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-544024"  
/clone\_1b="RPC1-11.2"  
30369..30381  
/note="Single clone region. Assembly confirmed by restriction digest data."  
BASE COUNT 25942 a 16986 c 16876 g 26032 t  
ORIGIN  
Query Match 13.7%; Score 21; DB 9; Length 85836;  
Best Local Similarity 100.0%; Pred. No. 1.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AL391556 95178 bp DNA linear HTG 10-JUL-2001  
 LOCUS AL391556  
 DEFINITION Homo sapiens chromosome 1 clone RP4-630J2 map q21.1-21.3, \*\*\*  
 SEQUENCING IN PROGRESS \*\*\*, 19 unordered pieces.  
 AL391556  
 VERSION AL391556.4 GI:10716368  
 KEYWORDS HTG: HTGS\_PHASE1; HTGS\_CANCELLED.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 95178)  
 McIay,K.  
 Direct Submission  
 Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonequest@sanger.ac.uk  
 On Oct 7, 2000 this sequence version replaced gi:10040155.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 ----- Project Information  
 Center project name: du630J2  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; 108752; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 56% of reads  
 Chemistry: Dye-terminator ET-amersham; 43% of reads  
 Consensus quality: 89533 bases at least Q40  
 Consensus quality: 91633 bases at least Q30  
 Consensus quality: 93378; sum-of-contigs  
 Insert size: 125478; 4.5% error; agarose-fp  
 Insert coverage: 3.24x in Q20 bases; sum-of-contigs  
 Quality coverage: 2.64x in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 19 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 \* 1 3063: contig of 3063 bp in length  
 \* 3064 3163: gap of 100 bp  
 \* 3164 6176: contig of 3013 bp in length  
 \* 6177 6276: gap of 100 bp  
 \* 6277 9333: contig of 3057 bp in length  
 \* 9334 9433: gap of 100 bp  
 \* 9434 18624: contig of 9191 bp in length  
 \* 18625 18724: gap of 100 bp  
 \* 18725 26968: contig of 8244 bp in length  
 \* 26969 27068: gap of 100 bp  
 \* 27069 37152: contig of 10084 bp in length  
 \* 37153 37252: gap of 100 bp  
 \* 37253 44300: contig of 7048 bp in length  
 \* 44301 44400: gap of 100 bp  
 \* 44401 46867: contig of 2467 bp in length  
 \* 46868 46967: gap of 100 bp  
 \* 46968 50594: contig of 3627 bp in length  
 \* 50595 50694: gap of 100 bp  
 \* 50695 53352: contig of 2658 bp in length  
 \* 53353 53452: gap of 100 bp  
 \* 53453 61187: contig of 7735 bp in length  
 \* 61188 61287: gap of 100 bp  
 \* 61288 65069: contig of 3782 bp in length  
 \* 65070 65169: gap of 100 bp  
 \* 65170 69334: contig of 4165 bp in length  
 \* 69335 69434: gap of 100 bp  
 \* 69435 77503: contig of 8069 bp in length

77504 77603: gap of 100 bp  
 \* 77604 79864: contig of 2261 bp in length  
 \* 79865 79964: gap of 100 bp  
 \* 79965 84018: contig of 4054 bp in length  
 \* 84019 84118: gap of 100 bp  
 \* 84119 86596: contig of 2478 bp in length  
 \* 86597 86696: gap of 100 bp  
 \* 86697 92135: contig of 5439 bp in length  
 \* 92136 92235: gap of 100 bp  
 \* 92236 95178: contig of 2943 bp in length.  
 FEATURES  
 source  
 1. 95178  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="q21.1-21.3"  
 /clone="RP4-630J2"  
 /clone\_1tb="RP4-4"  
 1. 3063  
 /note="assembly\_fragment:00100  
 fragment\_chain:1"  
 3164. 6176  
 /note="assembly\_fragment:00664  
 fragment\_chain:1"  
 6277. 9333  
 /note="assembly\_fragment:00660  
 fragment\_chain:1"  
 9434. 18624  
 /note="assembly\_fragment:00535  
 fragment\_chain:2"  
 18725. 26968  
 /note="assembly\_fragment:00354  
 fragment\_chain:2"  
 27069. 37152  
 /note="assembly\_fragment:00346  
 fragment\_chain:2"  
 37253. 44300  
 /note="assembly\_fragment:00798  
 fragment\_chain:3"  
 44401. 46867  
 /note="assembly\_fragment:00892  
 fragment\_chain:3"  
 46968. 50594  
 /note="assembly\_fragment:00511  
 fragment\_chain:3"  
 50695. 53352  
 /note="assembly\_fragment:00999  
 fragment\_chain:4"  
 53453. 61187  
 /note="assembly\_fragment:01015  
 fragment\_chain:4"  
 61288. 65069  
 /note="assembly\_fragment:00023"  
 65170. 69334  
 /note="assembly\_fragment:00187"  
 69435. 77503  
 /note="assembly\_fragment:00323"  
 77604. 79864  
 /note="assembly\_fragment:00464"  
 79965. 84018  
 /note="assembly\_fragment:00514"  
 84119. 86596  
 /note="assembly\_fragment:00677"  
 86697. 92135  
 /note="assembly\_fragment:00869"  
 92236. 95178  
 /note="assembly\_fragment:00941"  
 /note="assembly\_fragment:00941"  
 BASE COUNT 27407 a 18985 c 18958 g 28022 t 1806 others  
 ORIGIN  
 Query Match 13.7%; Score 21; DB 2; Length 95178;  
 Best Local Similarity 100.0%; Pred. No. 1.9;

```
Matches 21: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 111 TAAAGACCTTGCAGAAAAGA 131
|||||
Db 59055 TAAAGACCTTGCAGAAAAGA 59075

RESULT 13
AC016280 144979 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-20N19, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC016280
AC016280.3 GI:9123976
VERSION HTG: HTGS_PHASE0.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 144979)
Birten,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-20N19
Unpublished
2 (bases 1 to 144979)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Bara,N., Beckerly,R., Boguslavsky,L., Boukhgatter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Deavellano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Gallagan,J., Garday,S., Grant,G., Hagos,B., Harford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lien,C., Locken,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6730879.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W18R
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: 13944
Center clone name: 20_N19
-----
* NOTE: This record contains 138 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 872 971: gap of 100 bp
* 972 1881: contig of 910 bp in length
* 1882 1981: gap of 100 bp
* 1982 2869: contig of 888 bp in length
* 2870 2969: gap of 100 bp
* 2970 3880: contig of 911 bp in length
* 3881 3980: gap of 100 bp
* 3981 4853: contig of 873 bp in length
*
* 4854 4953: gap of 100 bp
* 4954 5827: contig of 874 bp in length
* 5828 5927: gap of 100 bp
* 5928 6786: contig of 859 bp in length
* 6787 6886: gap of 100 bp
* 6887 7820: contig of 934 bp in length
* 7821 7920: gap of 100 bp
* 7921 8850: contig of 930 bp in length
* 8851 8950: gap of 100 bp
* 8951 9824: contig of 874 bp in length
* 9825 9924: gap of 100 bp
* 9925 10845: contig of 921 bp in length
* 10846 10945: gap of 100 bp
* 10946 11851: contig of 906 bp in length
* 11852 11951: gap of 100 bp
* 11952 12829: contig of 878 bp in length
* 12830 12929: gap of 100 bp
* 12930 13796: contig of 867 bp in length
* 13797 13866: gap of 100 bp
* 13867 14796: contig of 900 bp in length
* 14797 14898: gap of 100 bp
* 14897 15813: contig of 917 bp in length
* 15814 15913: gap of 100 bp
* 15914 16813: contig of 902 bp in length
* 16816 16913: gap of 100 bp
* 16916 17827: contig of 912 bp in length
* 17828 17927: gap of 100 bp
* 17928 18809: contig of 882 bp in length
* 18810 18909: gap of 100 bp
* 18910 19788: contig of 879 bp in length
* 19789 19888: gap of 100 bp
* 19889 20799: contig of 911 bp in length
* 20800 20899: gap of 100 bp
* 20900 21829: contig of 930 bp in length
* 21830 21929: gap of 100 bp
* 21930 22837: contig of 908 bp in length
* 22838 22937: gap of 100 bp
* 22938 23812: contig of 875 bp in length
* 23813 23912: gap of 100 bp
* 23913 24797: contig of 885 bp in length
* 24798 24897: gap of 100 bp
* 24898 25757: contig of 860 bp in length
* 25758 25857: gap of 100 bp
* 25858 26754: contig of 897 bp in length
* 26755 26854: gap of 100 bp
* 26855 27766: contig of 912 bp in length
* 27767 27866: gap of 100 bp
* 27867 28790: contig of 924 bp in length
* 28791 28890: gap of 100 bp
* 28891 29789: contig of 899 bp in length
* 29780 29889: gap of 100 bp
* 29890 30786: contig of 897 bp in length
* 30787 30886: gap of 100 bp
* 30887 31757: contig of 871 bp in length
* 31758 31857: gap of 100 bp
* 31858 32736: contig of 879 bp in length
* 32737 32836: gap of 100 bp
* 32837 33691: contig of 855 bp in length
* 33692 33791: gap of 100 bp
* 33792 34669: contig of 878 bp in length
* 34670 34769: gap of 100 bp
* 34770 35679: contig of 910 bp in length
* 35680 35779: gap of 100 bp
* 35780 36669: contig of 890 bp in length
* 36670 36769: gap of 100 bp
* 36770 37674: contig of 905 bp in length
* 37675 37774: gap of 100 bp
* 37775 38676: contig of 902 bp in length
* 38677 38776: gap of 100 bp
* 38777 39679: contig of 903 bp in length
* 39680 39779: gap of 100 bp
* 39780 40654: gap of 875 bp in length
* 40655 40754: gap of 100 bp
```

```

* 40755 41619: contig of 865 bp in length
* 41620 41719: gap of 100 bp
* 41720 42559: contig of 840 bp in length
* 42560 42659: gap of 100 bp
* 42660 43575: contig of 916 bp in length
* 43576 43675: gap of 100 bp
* 43676 44570: contig of 895 bp in length
* 44571 44670: gap of 100 bp
* 44671 45577: contig of 907 bp in length
* 45578 45677: gap of 100 bp
* 45678 46590: contig of 913 bp in length
* 46591 46690: gap of 100 bp
* 46691 47556: contig of 866 bp in length
* 47557 47656: gap of 100 bp
* 47657 48546: contig of 890 bp in length
* 48547 48646: gap of 100 bp
* 48647 49517: contig of 871 bp in length
* 49518 49617: gap of 100 bp
* 49618 50511: contig of 894 bp in length
* 50512 50611: gap of 100 bp
* 50612 51543: contig of 932 bp in length
* 51544 51643: gap of 100 bp
* 51644 52559: contig of 916 bp in length
* 52560 52659: gap of 100 bp
* 52660 53547: contig of 888 bp in length
* 53548 53647: gap of 100 bp
* 53648 54528: contig of 881 bp in length
* 54529 54628: gap of 100 bp
* 54629 55500: contig of 872 bp in length
* 55501 55600: gap of 100 bp
* 55601 56503: contig of 903 bp in length
* 56504 56603: gap of 100 bp
* 56604 57517: contig of 914 bp in length
* 57518 57617: gap of 100 bp
* 57618 58500: contig of 883 bp in length
* 58501 58600: gap of 100 bp
* 58601 59484: contig of 884 bp in length
* 59485 59584: gap of 100 bp
* 59585 60487: contig of 903 bp in length
* 60488 60587: gap of 100 bp
* 60588 61508: contig of 921 bp in length
* 61509 61608: gap of 100 bp
* 61609 62533: contig of 925 bp in length
* 62534 62633: gap of 100 bp
* 62634 63476: contig of 843 bp in length
* 63477 63576: gap of 100 bp
* 63577 64458: contig of 882 bp in length
* 64459 64558: gap of 100 bp
* 64559 65442: contig of 884 bp in length
* 65443 65542: gap of 100 bp
* 65543 66455: contig of 913 bp in length
* 66456 66555: gap of 100 bp
* 66556 67462: contig of 907 bp in length
* 67463 67562: gap of 100 bp
* 67563 68550: contig of 988 bp in length
* 68551 68650: gap of 100 bp
* 68651 69609: contig of 959 bp in length
* 69610 69709: gap of 100 bp
* 69710 70779: contig of 1070 bp in length
* 70780 70879: gap of 100 bp
* 70880 71850: contig of 971 bp in length
* 71851 71950: gap of 100 bp
* 71951 72978: contig of 1028 bp in length

```

Query Match 13.7%: Score 21; DB 2: Length 144979;  
 Best Local Similarity 100.0%; Pred No. 1.8;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AAAGACCTTGCAAGAAAGAA 132  
 ||||||||||||||||||||  
 DB 136773 AAAGACCTTGCAAGAAAGAA 136793

```

RESULT 14
AC017081 149462 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-470J24 from 2, complete sequence.
AC017081
AC017081.8 GI:18093316
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 149462)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 149462)
Nguyen, C., Doebber, A. and Kozlowski, A.
The sequence of Homo sapiens BAC clone RP11-470J24
Unpublished (2001)
3 (bases 1 to 149462)
Waterston, R.H.
Direct Submission
Submitted (09-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 149462)
Waterston, R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jan 9, 2002 this sequence version replaced gi:14165368.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0470J24
-----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
 The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
 VECTOR: pBAC3.6  
 NEIGHBORING SEQUENCE INFORMATION:  
 The clone sequenced to the left is RP11-310K15, 2000 bp overlap; the clone sequenced to the right is RP11-156A1, 2000 bp overlap.

Actual start of this clone is at base position 190775 of RP11-310K15; actual end is at base position 28935 of RP11-156a1

The sequence between 81601 and 81703 is covered only by a PCR product of clone DNA. Unresolved tandem repeat regions exist between 81548 and 83183, 126196 and 127426.

**FEATURES**

Location/Qualifiers

**Source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-470J24"
/clone.lib="RP11-470J24"
/rpl_family="MIR"
581..621
/rpl_family="(TGGG)n"
965..1037
/rpl_family="GC_rich"
1127..1561
/notice="match to EST AA056210 (NID:g154531) zf51g03.r1"
1771..2127
/notice="match to EST AW614820 (NID:g7320006) hg80c08.x1"
2055..2195
/rpl_family="MERL_type"
2216..2419
/rpl_family="MERL_type"
2604..2844
/rpl_family="MIR"
2991..3324
/notice="match to EST R35058 (NID:g791959) yh86d09.r1"
3207..3788
/notice="match to EST AW452004 (NID:s6992780)"
3486..3527
/rpl_family="AT_rich"
3501..3745
/notice="match to EST AA054208 (NID:s154531) zf51g03.r1"
3981..4017
/rpl_family="T_rich"
4231..4273
/rpl_family="AT_rich"
5518..5847
/notice="match to EST R80994 (NID:g857275) y194c08.r1"
5606..5833
/rpl_family="AT_rich"
5722..5758
/rpl_family="AT_rich"
5802..5831
/rpl_family="(TTTTA)n"
5808..6116
/rpl_family="Alu"
6133..6244
/rpl_family="MERL_type"
6965..7000
/rpl_family="AT_rich"
7020..7094
/rpl_family="MERL_type"
7982..8006
/rpl_family="AT_rich"
8000..8436
/rpl_family="L2"
8437..8540
/rpl_family="L1"
8541..8696
/rpl_family="L2"
8722..9684
/rpl_family="L1"
9378..9422
/rpl_family="(TTG)n"
9685..9973
/rpl_family="Alu"
9974..10197

```

repeat_c_region	/rpt_family="L1" 10280., 10305 /rpt_family="PANC)n" 10320., 10341 /rpt_family="AT_rich" 10333., 10679 /rpt_family="L1" 10681., 10757 /rpt_family="L1" 10758., 11038 /rpt_family="Alu" 11040., 12011 /rpt_family="L1" 11966., 12011 11999., 12591 /rpt_family="L1" 12392., 12706 /rpt_family="Alu" 12707., 12952 /rpt_family="L1" 12866., 13140 /rpt_family="L1" 13133., 16061 /rpt_family="L1" 16063., 17107 /rpt_family="L1" 17108., 17161 /rpt_family="MeLR" 17163., 17468 /rpt_family="MeRL_type" 17471., 17675 /rpt_family="MeLR" 17677., 17835 /rpt_family="L1" 17909., 18144 /rpt_family="L1" 18177., 18501 /rpt_family="L1" 18519., 20393 /rpt_family="L1" 20490., 20906 /rpt_family="L1" 21258., 21365 /rpt_family="MeLR" 21353., 21393 /rpt_family="(?)n" 21366., 21684 /rpt_family="Alu" 21343., 21566 /rpt_family="AT_rich" 21683., 21813 /rpt_family="MeLR" 22160., 22640 /rpt_family="MeR2_type" 22876., 23228 /rpt_family="L1" 23687., 23991
-----------------	--

Query Match	13.7%;	Score 21;	DB 9;	Length 149462;
Best Local Similarity	100.0%;	Pred. No. 1.7;		
Matches 21;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	112	AAAGACCTTGCAAGAAAAGAA	132
Db	111213	AAAGACCTTGCAAGAAAAGAA	1112333

RESULT	15
LOCUS	AC024119
DEFINITION	AC024119 Homo sapiens chromosome 3, WORKING DRAFT SEQUENCE, 22 unordered places.
ACCESSION	AC024119
VERSION	AC024119.2 GI:10305055

KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 150436)  
Xu,S., Zhao,Y., Lin,W., Dong,H., Wan,M., Zhang,C., Gu,W., Tu,Y.,  
Jia,J., Wu,C., Lu,G., Zhong,M., Zhou,Y., Ren,S., Fu,G., Chen,Z. and  
Huang,M.  
TITLE Direct Submission  
JOURNAL Submitted (24-FEB-2000) Genomic Dept., Chinese National Human  
Genome Center at Shanghai, 315 Guo Shou Jing Road, Shanghai,  
Shanghai 201203, P. R. China  
On Sep 26, 2000 this sequence version replaced gi:7025692.  
COMMENT -----Genome Center-----  
Center:Beijing  
Center code:Beijing  
Website:http://hgclgtp.ac.cn  
http://www.genomics.org.cn  
Contact:hgc@lgtp.ac.cn  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 22 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
2766: contig of 2766 bp in length  
2767  
5770: contig of 3004 bp in length  
5771  
8549: contig of 2779 bp in length  
8550  
10693: contig of 2144 bp in length  
10694  
13973: contig of 3280 bp in length  
13974  
16980: contig of 3007 bp in length  
16981  
20419: contig of 3439 bp in length  
20420  
23275: contig of 2856 bp in length  
23276  
26837: contig of 3562 bp in length  
26838  
31026: contig of 4189 bp in length  
31027  
35204: contig of 4178 bp in length  
35205  
42487: contig of 7283 bp in length  
42488  
49419: contig of 6932 bp in length  
49420  
57524: contig of 8105 bp in length  
57525  
65794: contig of 8270 bp in length  
65795  
69733: contig of 3939 bp in length  
69734  
81835: contig of 12102 bp in length  
81836  
87228: contig of 5393 bp in length  
87229  
106006: contig of 18778 bp in length  
106007  
122564: contig of 16558 bp in length  
122565  
128482: contig of 5918 bp in length  
128483  
150436: contig of 21954 bp in length.  
FEATURES  
Source  
1. 150436  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="3"  
/clone\_lib="542323"  
BASE COUNT 45301 a 29770 c 31151 g 44136 t 78 others  
ORIGIN  
Query Match 13.7%; Score 21; DB 2; length 150436;  
Best Local Similarity 100.0%; Pred. No. 1.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 111 TAAAGACCTTGCAAGAAAGA 131  
|||||  
Db 91838 TAAAGACCTTGCAAGAAAGA 91858

Search completed: August 3, 2002, 01:43:18  
Job time: 40768 sec



Mon Aug 5 11:51:53 2002

us-08-973-363-5.oli.rge

Page 11



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 01:50:07 ; Search time 949.74 Seconds  
(without alignments)  
276.589 Million cell updates/sec

Title: US-08-973-363-5

Perfect score: 133

Sequence: 1 ATTTACTGATGATGCCAGCA.....TCGAAAGACTTACTGTGCA 153

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size: 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N\_Geneseq\_032802.\*

```
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	138	90.2	153	18	AAT42759
2	53	34.6	153	18	AAT42758
3	47	30.7	1316	18	AAT42754
4	38	24.8	153	18	AAT42757
5	38	24.8	6608	18	AAT42751
6	21	13.7	1409	19	AAV4574
7	21	13.7	1462	20	AAV56240
8	20	13.1	557	17	AAT11338
9	20	13.1	1157	21	AAC47521

10	20	13.1	1160	21	AAC34655	Arabidopsis thaliana
11	19	12.4	2580	21	AA298313	A. thaliana gene 1
12	19	12.4	2532	22	AA159877	Human polyomucleoti
13	18	11.8	44861	24	AA520000	DNA encoding pyrid
14	18	11.8	552	22	AA68219	Human lung tumour
15	18	11.8	1580	22	AA68219	Human homomylutamin
16	18	11.8	3225	19	AAV20467	Human c-myc oncoge
17	17	11.1	165	21	AAC09900	Human secreted pro
18	17	11.1	1788	21	AA56344	Escherichia coli f
19	17	11.1	2395	21	AAC43194	Arabidopsis thaliana
20	17	11.1	5647	22	AA199401	Human excretory re
21	17	11.1	5647	22	AA199401	Human kidney relat
22	16	10.5	136	15	AA724182	Human c-myc hammer
23	16	10.5	136	15	AA724182	5' end fragment of
24	16	10.5	196	18	AA789876	Staphylococcus aur
25	16	10.5	351	14	AAQ45230	Partial sequence o
26	16	10.5	385	22	AA67429	Novel human polynu
27	16	10.5	453	18	AA67429	Porcine somatotrop
28	16	10.5	537	22	ABA61160	Human foetal liver
29	16	10.5	537	22	AAK09455	Human brain expres
30	16	10.5	537	22	AAK35347	Human bone marrow
31	16	10.5	537	22	AA141060	Probe #9746 used t
32	16	10.5	594	22	AAK91625	Human CDNA 5'-end
33	16	10.5	594	22	AAK93548	Human CDNA clone r
34	16	10.5	630	14	AAQ47367	Sequence encoding
35	16	10.5	630	15	AAQ70199	DNA encoding novel
36	16	10.5	634	23	AA593365	Sequence encoding
37	16	10.5	681	14	AAQ47369	Sequence encoding
38	16	10.5	681	15	AAQ70201	Sequence encoding
39	16	10.5	690	14	AAQ47362	Sequence encoding
40	16	10.5	690	15	AAQ70209	Sequence encoding
41	16	10.5	690	15	AAQ70191	Sequence encoding
42	16	10.5	696	13	AAQ20237	Sequence encoding
43	16	10.5	699	13	AAQ20236	Encodes Val(1) to
44	16	10.5	699	14	AAQ47363	Encodes human beta
45	16	10.5	699	15	AAQ70192	Sequence encoding

## ALIGNMENTS

RESULT 1	
ID	AAT42759 standard; DNA: 153 BP.
XX	XX
AC	AAT42759:
XX	XX
DT	12-MAR-1997 (first entry)
XX	XX
DE	Great tit CHD-W gene fragment.
XX	XX
KW	Bird; sex determination; chromodomain-Helicase-DNA binding 1;
KM	CHD-1A; CHD-W; W chromosome; ss.
XX	XX
OS	Parus major.
XX	XX
PH	Key
FT	misc_difference 52..81
FT	/*tag=
FT	/note="bases 52-81 are a repeat of bases 22-51
FT	and are ignored in the translated amino
FT	acid sequence given in Fig 3"
XX	XX
PN	W09639505-A1.
XX	XX
PD	12-DEC-1996.
XX	XX
PF	05-JUN-1996; 96MO-GB01341.
XX	XX
PR	06-JUN-1995; 95GB-0011439.
XX	XX
RA	(ISIS-) ISIS INNOVATION LTD.
XX	XX



XX Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;  
SQ

Query Match 30.7%; Score 47; DB 18; Length 1316;  
Best Local Similarity 100.0%; Pred. No. 2,9e-14;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 82 CGTCGAGATTACCTCATTAATTAAGACCTTGCAAGAAA 128  
|||||  
Db 1189 cgtgagattacctatcaataatctgataaagacctgcaagaaa 1235

RESULT 4  
AAAT42757  
AAAT42757 standard; DNA; 153 BP.

AC AAAT42757;  
XX 12-MAR-1997 (first entry)  
XX Chick CHD-1A gene fragment.  
XX  
XX Bird: sex determination; chromodomain-Helicase-DNA binding 1;  
XX CHD-1A; CHD-W; W chromosome; ss.  
XX  
XX Gallus sp.  
XX  
XX Key Location/Qualifiers  
XX msc\_difference 52..81  
XX /\*tag= a  
XX /\*note= "bases 52-81 are a repeat of bases 22-51  
XX and are ignored in the translated amino  
XX acid sequence given in Fig 3"

PN MO9639505-A1.  
PD 12-DEC-1996.  
XX  
XX 05-JUN-1996; 96MO-GB01341.  
XX  
XX 06-JUN-1995; 95GB-0011439.  
XX  
XX (ISIS-) ISIS INNOVATION LTD.  
XX  
XX Griffiths R, Tiwari B;  
XX WPI: 1997-043127/04.  
XX P-PSDB: AAM08147.  
XX  
XX Avian chromodomain-helicase-DNA binding genes determine sex in  
XX birds - used for sex determin. and to control sex of progeny  
XX  
XX Claim 8; Fig 3; 76pp; English.  
XX  
XX Bases 3855-3977 (AAAT42756) of the mouse CHD-1 gene show homology  
XX to portions of the chicken CHD-1A (A - Avian) gene (AAAT42757).  
XX chicken CHD-W (W refers to the W chromosome) gene (see also AAAT42758)  
XX and the great tit CHD-W gene (see also AAAT42759). Translated amino  
XX acid sequences of this region are provided in AAM08146-49. The  
XX CHD-1A (see also AAAT42751) and CHD-W (see also AAAT42754-55) genes  
XX determine sex in birds and can be used to identify the sex of an  
XX embryo, fetus etc. and to manipulate the sex of progeny.  
XX  
XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

Query Match 24.8%; Score 38; DB 18; Length 153;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 91 TACCTCATTAATTAATTAAGACCTTGCAAGAAA 128  
|||||

Db 91 taccatcaataactgaataaagacctgcaagaaa 128

RESULT 5  
AAAT42751  
ID AAAT42751 standard; cDNA; 6608 BP.  
XX  
XX AAAT42751;  
XX  
XX 12-MAR-1997 (first entry)  
XX  
XX Chicken CHD-1A gene.  
XX  
XX Bird: sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
XX CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.  
XX  
XX Gallus sp.  
XX  
XX Key Location/Qualifiers  
XX CDS 228..5390  
XX /\*tag= a

PN MO9639505-A1.  
PD 12-DEC-1996.  
XX  
XX 05-JUN-1996; 96MO-GB01341.  
XX  
XX 06-JUN-1995; 95GB-0011439.  
XX  
XX (ISIS-) ISIS INNOVATION LTD.  
XX  
XX Griffiths R, Tiwari B;  
XX WPI: 1997-043127/04.  
XX  
XX Avian chromodomain-helicase-DNA binding genes determine sex in  
XX birds - used for sex determin. and to control sex of progeny  
XX  
XX Claim 1; Fig 5; 76pp; English.  
XX  
XX The chicken CHD-W gene (AAAT42754) acting alone or in conjunction with  
XX the closely related CHD-1A gene (AAAT42751) is suggested to initiate  
XX female development in birds. The sequence of CHD-1A was deduced  
XX from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
XX library using a great tit CHD-W sequence (see also AAAT42755) as probe.  
XX The CHD-1A (A - Avian) gene shows close identity to the mouse CHD-1  
XX gene (see also AAAT42756-57). It is located on an autosome or 2  
XX chromosome. Probes based on CHD-W and CHD-1A give a non-rare  
XX specific signal on hybridisation to genomic DNA of a non-rare  
XX bird and can be used for sex determin. of a bird. CHD-1A nucleic  
XX acids can also be used to control the sex of the progeny of a bird.

SQ Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;  
Query Match 24.8%; Score 38; DB 18; Length 6608;  
Best Local Similarity 100.0%; Pred. No. 1e-09;  
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 91 TACCTCATTAATTAATTAAGACCTTGCAAGAAA 128  
|||||  
Db 4140 taccatcaataactgaataaagacctgcaagaaa 4177

RESULT 6  
AAV49574  
ID AAV49574 standard; cDNA to mRNA; 1409 BP.  
XX  
XX AAV49574;  
XX  
XX 21-OCT-1998 (first entry)  
XX

DE	Human lymphoma cell line U937 clone HP10136 cDNA #1.
XX	
KW	Transmembrane domain; human; nutrition; cytokine; cell proliferation;
KW	differentiation; immune system; stimulator; suppressor; regulator;
KW	hematopoiesis; activin; inhibitor; chemotactic; chemoketic; receptor;
KW	haemostatic; thrombolytic; ligand; anti-inflammatory; tumour; ds.
XX	
OS	Homo sapiens.
XX	
EH	Key
FT	Location/Qualifiers
FT	82..729
FT	/*tag= a
FT	/product= "transmembrane domain containing protein"
XX	
PN	W09821328-A2.
XX	
PD	22-MAY-1998.
XX	
PF	07-NOV-1997; 97WO-JP04056.
XX	
PR	13-NOV-1996; 96JP-0301429.
XX	
PA	(PROT-) PROTEGENE INC.
XX	(SAGA ) SAGAMI CHEM RES CENTRE.
XX	
PI	Kato S, Kobayashi M, Sekine S, Yamaguchi T;
XX	
DR	WPI; 1998-297932/26.
XX	P-PSDB; AAM64546.
PT	Human protein having transmembrane domain - useful for, e.g.
PT	research and nutrition
XX	
PS	Claim 4; Page 153-155; 205pp; English.
XX	
CC	AAV49550-V49559 are cDNA sequences which encode human proteins containing
CC	a transmembrane domain. These proteins can be used for, e.g. research
CC	and nutrition, and may have cytokine and cell
CC	proliferation/differentiation, immune stimulating/suppressing,
CC	haematopoiesis regulating, tissue growth, activin/inhibin,
CC	chemotactic/chemoketic, haemostatic and thrombolytic, receptor/ligand,
CC	anti-inflammatory or tumour inhibition activity.
XX	
SO	Sequence 1409 BP; 377 A; 278 C; 325 G; 429 T; 0 other;
XX	
Query Match	13.7%; Score 21; DB 19; Length 1409;
Best Local Similarity	100.0%; Pred. No. 0.38;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
OY	111 TAAAGACCTTGCAGAAAGA 131
Db	842 taaagaccttgcagaagaa 862
XX	
RESULT 7	
AA556240	
ID	AA556240 standard; cDNA; 1462 BP.
XX	
AC	AA556240;
XX	
DT	16-JUL-1999 (first entry)
XX	
DE	Human CBFBA01 vesicle trafficking protein SEC22b gene.
XX	
KW	Human; vesicle trafficking protein; SEC22b; CBFBA01; cancer;
KW	autoimmune disease; diabetes mellitus; multiple sclerosis;
KW	diagnosis; ss.
XX	
OS	Homo sapiens.
XX	
PN	W09921988-A1.
XX	

PD	06-MAY-1999.
XX	
Pf	29-OCT-1997; 97WO-CN00115.
XX	
PR	29-OCT-1997; 97WO-CN00115.
XX	
PA	(USSH-) UNIV SHANGHAI SECOND MEDICAL.
XX	
PI	Wang Y, Wu J;
DR	WPI: 1999-312961/26.
XX	
DR	P-Psdb: AAY09512.
XX	
PF	CBFBA01 protein useful in treating cancer, autoimmune disease,
XX	diabetes mellitus and multiple sclerosis
PS	
XX	Claim 2; Page 21; 29pp; English.
CC	The present sequence encodes CBFBA01, which is a protein homologous to
CC	murine vesicle trafficking protein sec22b. CBFBA01 is a
CC	cytoplasmically-oriented integral membrane protein, located in the
CC	endoplasmic reticulum and golgi membranes. CBFBA01 polynucleotide (I),
CC	vectors containing (I) and recombinant host cells are useful for
CC	recombinant production of CBFBA01. Both (I), CBFBA01 and antibodies
CC	against CBFBA01 are useful as research reagents, for screening assays
CC	and in diagnostic assays, especially for cancer, autoimmune disease,
CC	diabetes mellitus and multiple sclerosis. Antagonists and agonists of
CC	CBFBA01 can be used to inhibit or enhance, respectively, the activity
CC	of CBFBA01 or expression of (I). Anti-CBFBA01 antibodies and CBFBA01
CC	or its fragments can be used in vaccines.
SQ	Sequence 1462 BP; 419 A; 286 C; 323 G; 434 T; 0 other;
Query Match	13.7%; Score 21; DB 20; Length 1462;
Best Local Similarity	100.0%; Pred. No. 0.38;
Matches 21; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	111 TAAAGACCTTCGACGAAGAAGA 131 
DB	826 taagacccttgcagaagaaga 846
RESULT 8	
ID	AAT11338
XX	AAT11338 standard; cDNA; 557 BP.
AC	AAT11338;
DT	05-SEP-1996 (first entry)
XX	
DE	AFT-1 interacting protein (partial ascorbate peroxidase) cDNA.
XX	
KM	Arabidopsis fourteen-three-three 1; AFT1; Cruciferae; regulation;
KW	transcription activator; gene expression modulator; crucifer;
KX	signal transduction; defence response; gene therapy; transgenic;
XX	plant; ss.
OS	Arabidopsis sp.
XX	
FH	Key Location/Qualifiers
FT	CDS 2..370
FT	/tag= a
FT	/product= partial ascorbate peroxidase
FT	/note= "plant defence related protein"
PN	EP693554-A1.
XX	
PD	24-JAN-1996.
XX	
PF	22-JUN-1995; 95EP-0109669.
XX	
PR	23-JUN-1994; 94US-0266451.

XX (GENO ) GEN HOSPITAL CORP.  
PA (MASS-) MASSACHUSETTS GEN HOSPITAL.  
XX  
PI Goodman HM, Zhang H;  
XX WPI: 1996-078601/09.  
DR P-PSDB: MAR89746.  
XX  
PT Recombinant transcriptional activator APT-1 - isolated from  
PT Arabidopsis and useful for modifying gene expression and modulating  
PT plant defence responses to pathogens  
XX  
PS Disclosure; Page 21; 47pp; English.  
XX  
CC AAT11338 encodes a partial ascorbate peroxidase, a plant defence-  
CC related protein (PDRP). AAT11338 and other cDNA molecules encoding  
CC PDRPs were isolated using a LexA/APT1 33-194 fusion protein, i.e.  
CC APT1 (Arabidopsis 14-3-3-1) protein residues 33-194 fused to LexA,  
CC a DNA binding protein, in an interaction trap assay. APT1 is a  
CC transcriptional activator protein and may be used to enhance,  
CC control or modify plant gene expression, e.g. in the regulation of  
CC plant storage components (esp. napin, legumin or phaseolin). APT1  
CC can also be used to modulate signal transduction events involved  
CC in plant defence responses to pathogens such as fungi, nematodes,  
CC bacteria and viruses.  
XX  
SQ Sequence 557 BP; 164 A; 110 C; 133 G; 150 T; 0 other;  
  
Query Match 13.1%; Score 20; DB 17; Length 557;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 31 CAGGCAAGACGTCGACAC 50  
|||||  
Db 255 caggaagaagcagctgcagac 274  
  
RESULT 9  
AAC47521  
ID AAC47521 standard; DNA; 1157 BP.  
XX  
AC AAC47521;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54126.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP103405-A2.  
XX  
XX  
XX PD 06-SEP-2000.  
XX  
XX  
XX PE 25-FEB-2000; 2000EP-0301439.  
XX  
XX PR 25-FEB-1999; 99US-0121825.  
XX PR 05-MAR-1999; 99US-0123180.  
XX PR 09-MAR-1999; 99US-0123548.  
XX PR 23-MAR-1999; 99US-0125788.  
XX PR 25-MAR-1999; 99US-0126264.  
XX PR 29-MAR-1999; 99US-0126785.  
XX PR 01-APR-1999; 99US-0127463.  
XX PR 06-APR-1999; 99US-0128234.  
XX PR 08-APR-1999; 99US-0128714.  
XX PR 16-APR-1999; 99US-0128843.  
XX PR 19-APR-1999; 99US-0130077.  
XX PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134270.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142055.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145313.  
PR 27-JUL-1999; 99US-0145318.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147760.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150684.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 13.1%; Score 20; DB 21; Length 1157;  
Best Local Similarity 100.0%; Pred. No. 1.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 31 CAGCAAGCAGTTCGAC 50  
Db 860 caggcaagcagltgcagac 879  
  
RESULT 10  
AAC34655  
ID AAC34655 standard; DNA; 1160 BP.  
XX AC AAC34655;  
XX DT 17-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 7404.  
XX KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
XX metabolic pathway; promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-0301439.  
XX PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.



PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132853.  
PR 14-MAY-1999; 99US-0134226.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137232.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138004.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139432.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140921.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142380.  
PR 09-JUL-1999; 99US-0142803.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147923.  
PR 11-AUG-1999; 99US-0148317.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148365.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 13-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 18-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.

```
PR 21-OCT-1999; 990S-0160814.  
PR 21-OCT-1999; 990S-0160815.  
PR 22-OCT-1999; 990S-0160980.  
PR 22-OCT-1999; 990S-0160981.  
PR 22-OCT-1999; 990S-0160989.  
PR 25-OCT-1999; 990S-0161404.  
PR 25-OCT-1999; 990S-0161405.  
PR 25-OCT-1999; 990S-0161406.  
PR 26-OCT-1999; 990S-0161359.  
PR 26-OCT-1999; 990S-0161360.  
PR 26-OCT-1999; 990S-0161361.  
PR 28-OCT-1999; 990S-0161920.  
PR 28-OCT-1999; 990S-0161992.  
PR 28-OCT-1999; 990S-0161993.  
PR 29-OCT-1999; 990S-0162142.
```

```
Query Match 13.1%; Score 20; DB 21; Length 1160;  
Best Local Similarity 100.0%; Pred. No. 1.2; Mismatches 0; Indels 0; Gaps 0;  
Matches 20; Conservative 0;
```

```
QY 31 CAGGCAAGCAGTTGCAGAC 50  
|||||  
Db 863 cagcgcaagcagctgcagac 882
```

```
RESULT 11  
AA298313  
ID AA298313 standard; DNA; 2580 BP.
```

```
AC AA298313;
```

```
DT 14-JUN-2000 (first entry)
```

```
DE A. thaliana gene involved in environmental stress tolerance.
```

```
XX Environmental stress; plant; transgenic plant; anaerobic; flooding; cold;  
KM dehydration; drought; heat stress; salinity; osmotolerance; ds.
```

```
XX Arabidopsis thaliana.
```

```
XX WO200008187-A2.
```

```
XX 17-FEB-2000.
```

```
XX 04-AUG-1999; 99WO-EP05652.
```

```
XX 04-AUG-1998; 98EP-0202634.
```

```
XX (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
```

```
XX Lee JH, Verbruggen N;
```

```
XX WPI: 2000-205726/18.
```

```
XX P-PSDB; AAT77933.
```

```
PT Isolation of polynucleic acids useful for producing transgenic plant by  
isolating genes involved in tolerance to environmental stress
```

```
PS Claim 4; Page 112-115; 312pp; English.
```

The invention relates to isolation of coding sequences and/or genes involved in tolerance to environmental stress in plants. The sequences (AA298305-298365) are useful for producing a transgenic plant having enhanced tolerance or resistance to environmental stress conditions such as anaerobic, flooding, cold, dehydration, drought, heat stress or salinity. This is useful for producing improved yield, growth, development and productivity under environmental stress conditions, and also provides growth of crops in areas where they cannot grow without the induced osmotolerance. Sequences AA298305-365 represent polynucleotide sequences from A. thaliana that are involved in environmental stress tolerance.

Sequence 2580 BP; 677 A; 459 C; 544 G; 900 T; 0 other;

```
Query Match 13.1%; Score 20; DB 21; Length 2580;  
Best Local Similarity 100.0%; Pred. No. 1.2; Mismatches 0; Indels 0; Gaps 0;  
Matches 20; Conservative 0;
```

```
QY 31 CAGGCAAGCAGTTGCAGAC 50  
|||||  
Db 2449 cagcgcaagcagctgcagac 2468
```

```
RESULT 12
```

```
AA158977  
ID AA158977 standard; cDNA; 2532 BP.
```

```
AC AA158977;
```

```
DT 22-OCT-2001 (first entry)
```

```
DE Human polynucleotide SEQ ID NO 1180.
```

Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.

Homo sapiens.

WO200153312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;

Zhao QA, Zhou P, Goodrich R, Dirmanac RT;

WPI: 2001-442253/47.

P-PSDB; AAM39821.

Novel nucleic acids and polypeptides, useful for treating disorders

such as central nervous system injuries -

Claim 1; SEQ ID NO 1180; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAM38642-AAM42213) with neurotropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.



```

FT      variation      /standard_name="Single nucleotide polymorphism"
FT      /replace(40521,T)
FT      /tag= al
FT      /standard_name="Single nucleotide polymorphism"
FT      /replace(40563..40748
FT      /tag= am
FT      /number= 10
FT      /replace(40749..40861
FT      /tag= an
FT      /number= 11
FT      /replace(40841..A)
FT      /tag= ao
FT      /standard_name="Single nucleotide polymorphism"
FT      /replace(40875,C)
FT      /tag= ap
FT      /standard_name="Single nucleotide polymorphism"
FT      /
FT      WO200190125-A2.
FT      29-NOV-2001.
FT      24-MAY-2001; 2001WO-US16909.
FT      24-MAY-2000; 2000US-206664P.
FT      (GENA-) GENAISSANCE PHARM INC.
FT      Chew A, Duda A, Koshy B;
FT      WPI: 2002-106169/14.
FT      P-PSDB; AAU11951.
FT      Isolated human pyridoxal (pyridoxine, vitamin B6) kinase polyNTs,
FT      useful for therapeutic purposes, for studying the expression and
FT      function of the polyNT, and for expressing pyridoxal protein -
FT      Disclosure: Fig 1; 135pp; English.
FT      The invention describes an isolated human pyridoxal (pyridoxine, vitamin
FT      B6) kinase, (PDXK) polynucleotide. The polynucleotide is useful in
FT      studying the expression and function of PDXK, and in expressing PDXK
FT      protein for use in screening for candidate drugs to treat PDXK related
FT      diseases and for therapeutic purposes. A transgenic animal is useful for
FT      studying expression of the PDXK isogenes in vivo, for in vivo screening
FT      and testing of drugs targeted against PDXK protein, and for testing the
FT      efficacy of therapeutic agents and compounds for autoimmune polyglanular
FT      disease type 1. The polypeptide is useful for studying the effect of the
FT      variation on the biological activity of PDXK and the binding affinity of
FT      candidate drugs targeting PDXK for the treatment of autoimmune
FT      polyglanular disease type 1. Genotyping and haplotyping is useful for
FT      improving the efficacy and reliability of several steps in the discovery
FT      and development of drugs for treating diseases associated with PDXK
FT      activity, e.g., autoimmune polyglanular disease type 1, to validate PDXK
FT      as a candidate agent for treating a specific condition or disease
FT      predicted to be associated with PDXK activity, and in the design of
FT      clinical trials of candidate drugs. This sequence (located on chromosome
FT      21q22.3) encodes human pyridoxal (pyridoxine, vitamin B6) kinase (PDXK)
FT      and forms the reference sequence on which the isoforms AAU11952 and
FT      AAU11953 are based, described in the method of the invention.
FT      Sequence 44861 BP; 9041 A; 12297 C; 13398 G; 10125 T; 0 other:

```

XX	AAAF68219	
XX	AAAF68219 standard; cDNA; 552 BP.	
AC	AAAF68219;	
XX		
DT	12-APR-2001 (first entry)	
XX		
DE	Human lung tumour protein related nucleotide sequence SFG ID NO:137.	
XX		
KW	Human: lung cancer; lung tumour; lung tumour protein; gene therapy;	
KW	lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;	
KW	cytostatic; antisense inhibition; ss.	
XX		
OS	Homo sapiens.	
XX		
PD	NO200100828-A2.	
XX		
PN	04-JAN-2001.	
XX		
PF	30-JUN-2000; 2000WO-US18061.	
XX		
PR	30-JUN-1999; 99US-0346492.	
PR	15-OCT-1999; 99US-0419356.	
PR	17-DEC-1999; 99US-0466867.	
PR	30-DEC-1999; 99US-0476300.	
PR	06-MAR-2000; 2000US-0519642.	
PR	22-MAR-2000; 2000US-0533077.	
PR	10-APR-2000; 2000US-0546259.	
PR	27-APR-2000; 2000US-0560406.	
PR	05-JUN-2000; 2000US-0589184.	
XX		
PA	(CORI-) CORIXA CORP.	
XX		
PI	Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;	
PI	Retter MW, Mannion J;	
XX		
DR	WPI; 2001-071488/08.	
XX		
XX		
PS	Example 1; Page 193; 436pp; English.	
XX		
CC	The present invention describes immunogenic portions of lung tumour-	
CC	associated proteins (I) and the nucleic acids (NAs) that encode them.	
CC	(I) have cytostatic activity and can be used in gene therapy, antisense	
CC	inhibition and in vaccines. The NAs and the lung tumour-associated	
CC	proteins they encode may be used in the prevention, treatment and	
CC	diagnosis of diseases associated with their inappropriate expression,	
CC	especially lung cancers. For example, the NAs may be administered to	
CC	treat diseases by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of the protein by expressing inactive proteins	
CC	or to supplement the patients own production of (I). Additionally, the	
CC	NAs may be used to produce the lung-tumour associated protein, according	
CC	to standard recombinant DNA methodology. Conversely, antisense NA	
CC	molecules may be administered to down regulate protein expression by	
CC	binding with the cells own genes and preventing their expression. The NA	
CC	and complementary sequences may also be used as DNA probes in diagnostic	
CC	assays to detect and quantitate the presence of similar NA sequences in	
CC	samples, and hence which patients may be in need of treatment for lung	
CC	cancer. The (I) may be used as antigens in the production of antibodies	
CC	and in assays to identify modulators (agonists and antagonists) of the	
CC	expression and activity of the protein. AAF68083 to AAF68878 and	
CC	AA676848 to AAF76678 represent human lung tumour protein related	
CC	nucleotide and protein sequences which are used in the exemplification	
CC	of the present invention.	
XX		
XX	Sequence 552 BP; 220 A; 118 C; 69 G; 144 T; 1 other;	

```
Query Match      12.4%: Score 19: DB 24: Length 44861:
Best Local Similarity 100.0%: Pred. No. 3.8:
Matches 19: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY    50 CCAGGAACCAACAGGCGCAA 68
       |||iiiiiii|||iiiii
Db    1832 ccaagaaccacagcgcaa 1850
```

Query Match	11.8%;	Score 18;	DB 22;	Length 552;
Best Local Similarity	100.0%;	Pred. No. 13;		
Matches	18;	Conservative	0;	Mismatches
			0;	Indels
				Gaps
				0;

OY 97 ATTAATTACTGAATAA 114  
 |||||  
 DB 492 attaatcactgataaa 509

RESULT 15

AAAF89129  
 ID AAF89129 standard; DNA; 1580 BP.

AC AAF89129;

DT 13-JUL-2001 (first entry)

DE Human homoglutamine-rich factor 56 coding sequence.

KW Human: homoglutamine-rich factor 56; HQRF: cancer; angiodysplasia;  
 KW nervous system disease; immune disease; phlogosis; ds.

OS Homo sapiens.

Key Location/Qualifiers  
 CDS 27..1547  
 /tag=a  
 /product="homoglutamine-rich factor 56"

WO200127148-A1.

PD 19-APR-2001.

PF 08-OCT-2000; 2000WO-CN00310.

PR 10-OCT-1999; 99CN-0116946.

PA (SHAN-) SHANGHAI BIO DOOR GENE TECHNOLOGY LTD.

PI Mao Y, Xie Y;

DR WPI; 2001-281976/29.

DR P-PSDB; AAB83299.

PT Homo (human) glutamine(Q)-rich factor 56 applicable in diagnosis and  
 PT treatment of cancer, angiodysplasia, nervous system diseases, immune  
 PT diseases and phlogosis

PS Claim 6: Page 21-22: 29pp; Chinese.

CC The present invention provides the protein and coding sequences of human  
 CC homoglutamine-rich factor 56 (HQRF). This is useful in the treatment of  
 CC cancer, angiodysplasia, nervous system diseases, immune diseases and  
 CC phlogosis. The present sequence is the homoglutamine-rich factor 56  
 CC coding sequence.

SO Sequence 1580 BP; 463 A; 444 C; 346 G; 327 T; 0 other:

Query Match 11.8%; Score 18; DB 22; Length 1580;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 20 ATAGAAACCCACAGGCAA 37  
 |||||  
 DB 237 ataagaaccacagcgaa 254

Search completed: August 3, 2002, 01:50:11  
 Job time: 32196 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:37 ; Search time 7016.61 Seconds  
(without alignments) 294.307 Million cell updates/sec

Title: US-08-973-363-5

Sequence: 1 ATTTACTGATGATGACCCAGCA.....TGCAAAAGACTTACTGTGCA 153

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estdb:\*  
2: em\_estdb:\*  
3: em\_estdb:\*  
4: em\_estdb:\*  
5: em\_estdb:\*  
6: em\_estdb:\*  
7: em\_estdb:\*  
8: em\_estdb:\*  
9: em\_estdb:\*  
10: em\_estdb:\*  
11: em\_estdb:\*  
12: em\_estdb:\*  
13: em\_estdb:\*  
14: em\_estdb:\*  
15: em\_estdb:\*  
16: em\_estdb:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	14.4	167	10	BE842996
C 2	21	13.7	314	9	AA354813
C 3	21	13.7	322	10	BE289372
C 4	21	13.7	355	9	AI902726
C 5	21	13.7	438	9	AA234796
C 6	21	13.7	461	9	AA456385
C 7	21	13.7	474	10	BG66575
C 8	21	13.7	488	10	BG573124
C 9	21	13.7	489	10	BG573124
C 10	21	13.7	492	9	BE198540
C 11	21	13.7	515	9	AW579514
C 12	21	13.7	560	10	BM228157
C 13	21	13.7	589	9	AI412929
C 14	21	13.7	594	9	AM993220
C 15	21	13.7	597	10	BE168086
C 16	21	13.7	616	9	AM992861
C 17	21	13.7	616	10	BE690915

C 18	21	13.7	624	9	AM992921
C 19	21	13.7	628	9	AM958448
C 20	21	13.7	631	10	BM229259
C 21	21	13.7	643	9	AM992919
C 22	21	13.7	643	9	AM993201
C 23	21	13.7	647	9	AM993283
C 24	21	13.7	658	9	AM993467
C 25	21	13.7	675	9	AV232349
C 26	21	13.7	684	9	AV599376
C 27	21	13.7	694	9	AM993481
C 28	21	13.7	698	9	AM993471
C 29	21	13.7	706	9	AM993268
C 30	21	13.7	778	10	BI852172
C 31	21	13.7	793	10	BG182192
C 32	21	13.7	874	10	BG426048
C 33	21	13.7	924	10	BG180912
C 34	21	13.7	982	10	BI105781
C 35	21	13.7	1358	11	AK010363
C 36	21	13.7	1394	11	AK017237
C 37	21	13.7	1462	11	AK047442
C 38	20	13.1	294	9	AV531270
C 39	20	13.1	337	9	AM996787
C 40	20	13.1	366	9	AV566736
C 41	20	13.1	379	9	AV521243
C 42	20	13.1	411	10	Z34197
C 43	20	13.1	430	9	AL601246
C 44	20	13.1	502	9	AV518366
C 45	20	13.1	502	9	AV521265

## ALIGNMENTS

RESULT 1  
LOCUS BE842996/c  
DEFINITION MR4-ST0240-240700-013-e10 ST0240 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BE842996  
VERSION MR4-ST0240-240700-013-e10 ST0240  
KEYWORDS EST  
SOURCE BE842996.1 GI:10275374  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 167)  
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldmann,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL  
MEDLINE  
COMMENT  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Pudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=612-MR4-ST0240-240700-013-e10-ft3-2000-07-24-ft4-1)  
Seq primer: puc 18 forward  
High quality sequence start: 26  
High quality sequence stop: 167.  
Location/Qualifiers  
FEATURES  
source  
1..167  
/organism="Homo sapiens"

```

/db_xref="taxon:9606"
/clone_lib="ST0240"
/dev_stage="Adult"
/note="Organ: stomach; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      38 a      42 c      29 g      58 t
ORIGIN
Query Match      14.4%; Score 22; DB 10; Length 167;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 111 TAAAGACCTTGCAAGAAAGA 132
|||||
Db 46 TAAAGACCTTGCAAGAAAGA 25

RESULT 2
AA354813 314 bp mRNA linear EST 21-APR-1997
LOCUS EST63335 Jurkat T-cells V Homo sapiens cDNA 5' end, mRNA sequence.
DEFINITION AA354813
ACCESSION AA354813
VERSION AA354813.1 GI:2007319
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 314)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
,C.J., Lee,N.H., Kirness,E.F., Weinstein,K.G., Gocayne,J.D., White
,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Well,C., Clayton,R.A.,
Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
Gneltun,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Uteback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W.,
Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
Kunsch,C., Hungjun,U., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
Other_ESTs: THC180658
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tdb/ngi/ngi.html)
Seq primer: M13 Reverse.
FEATURES
source
1..314
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="ATCC (inhost):155260"
/db_xref="taxon:9606"

```

```

/clone_lib="Jurkat T-cells V"
/cell_type="T-Lymphocyte"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT      90 a      62 c      68 g      94 t
ORIGIN
Query Match      13.7%; Score 21; DB 9; Length 314;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 111 TAAAGACCTTGCAAGAAAGA 131
|||||
Db 61 TAAAGACCTTGCAAGAAAGA 81

RESULT 3
BF289372 322 bp mRNA linear EST 28-NOV-2000
LOCUS EST453963 Rat Gene Index, normalized rat, Rattus norvegicus cDNA
DEFINITION BF289372
ACCESSION BF289372
VERSION BF289372.1 GI:11220442
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 322)
Malek,R.L., Cho,J., Lee,Y., Karamycheva,S., Parvizl,B., Perlea,G.,
Sultana,R., Tsai,J., White,J., Quackenbush,J. and Lee,N.H.
Generation of ESTs from Normalized Rat Embryo, Bento Soares
Unpublished (2000)
CONTACT: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
This clone is available through the ATCC, contact the ATCC
tel#703-365-2700 for further information.
Location/Qualifiers
1..322
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone_lib="R61HE60"
/clone_lib="Rat Gene Index, normalized rat, Rattus
norvegicus cDNA"
/tissue_type="mixed tissue"
/lab_host="DH5-alpha"
/note="Vector: pT37Pac; Site_1: EcoRI; Site_2: NotI;
Combination of ROY, RBR, RRI, RLI, RPL, RUU, REM, RNU, RSP
, RHE, RPC, RPN"
BASE COUNT      77 a      71 c      80 g      94 t
ORIGIN
Query Match      13.7%; Score 21; DB 10; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 111 TAAAGACCTTGCAAGAAAGA 131
|||||
Db 18 TAAAGACCTTGCAAGAAAGA 38

RESULT 4
A1902726 355 bp mRNA linear EST 30-MAR-2000
LOCUS A1902726
DEFINITION QV-BT011-101198-028 BT011 Homo sapiens cDNA, mRNA sequence.
ACCESSION A1902726

```



```

VERSION      A1902726.1  GI:6493113
KEYWORDS
SOURCE       human
ORGANISM     Homo sapiens
              Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE    1 (bases 1 to 355)
AUTHORS      Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R.,
              Nagai M.A., da Silva M. Jr., Zago M.A., Bordin S., Costa F.F.,
              Goldman G.H., Carvalho A.F., Matsukuma A., Bais G.S., Simpson D.H.,
              Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V., O'Hare
              M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and
              Simpson A.J.
TITLE        Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
JOURNAL      Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
COMMENT      Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              Project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/seq/gethtml.pl?tl=QVet2-QV-BT011-028.html
              &t3=101198&t4=1)
              Seq primer: puc 18 forward.
FEATURES
  source
    1..355
       /organism="Homo sapiens"
       /db_xref="taxon:9606"
       /clone_lib="BT011"
       /sex="Female"
       /dev_stage="Adult"
       /note="Organ: breast; Vector: puc18; Site.1: SmaI; Site.2:
       SmaI; A mini-library was made by cloning products derived
       from ORESTES PCR (U.S. Patents Patent Application No. 196
       716 - Ludwig Institute for Cancer Research) profiles
       into the pUC 18 vector. Reverse transcription of tissue
       mRNA and cDNA amplification were performed under low
       stringency conditions."
BASE COUNT   107 a 66 c 80 g 102 t
ORIGIN
Query Match      13.7%; Score 21; DB 9; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 111 TAAAGACCTTCGACGAAAAGA 131
      ||||||||||||||||||
DB 116 TAAAGACCTTCGACGAAAAGA 136

RESULT 5
LOCUS       AA234796             438 bp      mRNA      linear      EST 06-AUG-1997
DEFINITION  z57805.r1 Soares_NhMMPu_S1 Homo sapiens cDNA clone IMAGE:665512
VERSION     AA234796
KEYWORDS    5', mRNA sequence.
SOURCE      AA234796.1  GI:1859289
              EST.
              human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE    1 (bases 1 to 438)
AUTHORS      Hallier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S.,
              Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J., Moore B.,
              Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J., Moore B.,
              Schellenberg K., Steptoe M., Tan F., Theising B., White Y., Wylie
              , Schellenberg K., Steptoe M., Tan F., Theising B., White Y., Wylie

```

```

TITLE        'T., Waterston R. and Wilson R.
              WashU-Merck EST Project 1997
              Unpublished (1997)
JOURNAL
COMMENT      Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available royalty-free through LNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Insert length: 875 Std Error: 0.00
              Seq primer: -28m3 rev2 ET from Amersham
              High quality sequence stop: 407.
FEATURES
  source
    1..438
       /organism="Homo sapiens"
       /db_xref="GDB:3563479"
       /db_xref="taxon:9606"
       /clone_lib="IMAGE:669512"
       /clone_lib="Soares_NhMMPu_S1"
       /tissue_type="Pooled human melanocyte, fetal heart, and
       pregnant uterus"
       /lab_host="DH10B"
       /note="Organ: mixed (see below); Vector: pTT73D-Pac
       (Pharmacia) with a modified polylinker; Site.1: Not I;
       Site.2: Eco RI; Equal amounts of plasmid DNA from three
       normalized libraries (melanocyte 2NDHM, pregnant uterus
       NBHPU, and fetal heart NBH19W) were mixed, and ss circles
       were made in vitro. Following HAP purification, this DNA
       was used as tracer in a subtractive hybridization
       reaction. The driver was PCR-amplified cDNAs from pools of
       5,000 clones made from the same 3 libraries. The pools
       consisted of 1 M.A.G.E. clones 260232-265223,
       340488-345479, and 484488-489479."
BASE COUNT     130 a 83 c 97 g 128 t
ORIGIN
Query Match      13.7%; Score 21; DB 9; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 111 TAAAGACCTTCGACGAAAAGA 131
      ||||||||||||||||||
DB 154 TAAAGACCTTCGACGAAAAGA 174

RESULT 6
LOCUS       AA456385             461 bp      mRNA      linear      EST 06-JUN-1997
DEFINITION  aa44f05.r1 Soares_NhMMPu_S1 Homo sapiens cDNA clone IMAGE:813249
VERSION     AA456385
KEYWORDS    5', mRNA sequence.
SOURCE      AA456385.1  GI:2178961
              EST.
              human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE    1 (bases 1 to 461)
AUTHORS      Hallier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S.,
              Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J., Moore B.,
              Schellenberg K., Steptoe M., Tan F., Theising B., White Y., Wylie
              , T., Waterston R. and Wilson R.
              WashU-Merck EST Project 1997
              Unpublished (1997)
              Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu

```

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -28m3 rev2 ET from Amer sham  
 High quality sequence stop: 446.

## FEATURES

source

Location/Qualifiers  
 1..461  
 /organism="Homo sapiens"  
 /db\_xref="GDB:6044159"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:813249"  
 /clone\_lib="Soares\_NHMPU-S1"  
 /tissue\_type="Pooled human melanocyte, fetal heart, and pregnant uterus"  
 /lab\_host="DH10B"  
 /note="Organ: mixed (see below); Vector: PT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI. Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NDM, pregnant uterus NBHPU, and fetal heart NBH19W) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

## BASE COUNT

132 a 85 c 103 g 141 t

## ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 461;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131  
 Db 155 TAAAGACCTTGCAAGAAAGA 175

## RESULT 7

LOCUS BG666575 474 bp mRNA linear EST 30-APR-2001  
 DEFINITION DRABNG04 Rat DRG Library Rattus norvegicus CDNA clone DRABNG04 5',  
 mRNA sequence.

ACCESSION BG666575  
 VERSION BG666575.1 GI:13888497

KEYWORDS EST.  
 SOURCE Norway rat.

ORGANISM Rattus norvegicus

Rattus norvegicus; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 474)  
 Xiao,H.S., Han,Z.G., Zhang,F.X., Huang,Q.H., Lu,Y.J., Bao,L., Fu,G.,  
 Guo,C., Yan,Q., Jin,S.X., Zhu,Z.D., Xu,X.R., Li,N.G., Chen,Z. and  
 Zhang,X.

AUTHORS  
 TITLE Distinct gene expression profiles of rat dorsal root ganglion  
 induced by peripheral nerve axotomy

JOURNAL  
 COMMENT Unpublished (2001)  
 Contact: Zhang Xu  
 Laboratory of Sensory System  
 Institute of Neuroscience  
 320 Yue Yang Road, Shanghai 200031, P.R.China  
 Tel: 86-21-64748700-121  
 Fax: 86-21-64713446

Email: xu.zhang@ion.ac.cn  
 This clone is also available at Chinese National Human Genome  
 Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,  
 Pudong New Area, P.R.China. Please contact with Zhang Xu  
 (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)

PCR Primers  
 FORWARD: T7  
 BACKWARD: T3

Seq primer: T3

FEATURES  
 source POLYA=NO.  
 Location/Qualifiers  
 1..474

/organism="Rattus norvegicus"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="DRABNG04"  
 /clone\_lib="Rat DRG Library"  
 /sex="male"  
 /tissue\_type="dorsal root ganglion"  
 /dev\_stage="adult"

BASE COUNT 110 a 116 c 125 g 123 t

## ORIGIN

Query Match 13.7%; Score 21; DB 10; Length 474;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131  
 Db 203 TAAAGACCTTGCAAGAAAGA 223

## RESULT 8

LOCUS BG573124 488 bp mRNA linear EST 10-APR-2001  
 DEFINITION 602594129F1 NIH\_MGC\_79 Homo sapiens CDNA clone IMAGE:4721382 5',  
 mRNA sequence.

ACCESSION BG573124  
 VERSION BG573124.1 GI:13580777

KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens

Homo sapiens; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 488)

REFERENCE NIH-MGC http://mgs.nci.nih.gov/.  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Clontech Laboratories, Inc.  
 CDNA Library Preparation: Clontech Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: L1CM1577 row: j column: 07  
 High quality sequence stop: 475.

## FEATURES

source

Location/Qualifiers  
 1..488

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4721382"  
 /clone\_lib="NIH\_MGC\_79"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Placenta; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfiI (ggcgcttcggc); Site\_2: SfiI (ggcattatgccc  
 ); 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGCGCATATGCGC-3' and 3' adaptor  
 sequence: 5'-ATTCGAGAGCGCGCGCGCATG-dT(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
 library."

BASE COUNT 136 a 88 c 108 g 156 t

## ORIGIN

Query Match 13.7%; Score 21; DB 10; Length 488;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTCAGAAAGA 131  
 |||||||  
 DB 176 TAAAGACCTTCAGAAAGA 196

RESULT 9  
 AM464549 489 bp mRNA linear EST 24-FEB-2000  
 LOCUS BP230016A10E9 Soares normalized bovine placenta Bos taurus CDNA  
 DEFINITION clone BP230016A10E9 5', mRNA sequence.  
 ACCESSION AM464549.1 GI:7034717  
 VERSION EST.  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM Bos taurus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 Bovidae; Bovinae; Bos.  
 REFERENCE 1 (bases 1 to 489)  
 AUTHORS Lewin,H.A., Soares,M.B., Rebelz,M., Pardinas,J., Liu,L. and Larson,J.H.  
 TITLE Bovine ESTs  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Lewin, H. A.  
 W. M. Keck Center for Comparative and Functional Genomics  
 University of Illinois at Urbana-Champaign  
 340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL  
 61801, USA  
 Tel: 217 333 5998  
 Fax: 217 244 5617  
 Email: h-lewin@uiuc.edu  
 Funding for cattle EST sequencing was provided by the USDA National  
 Research Initiative, Animal Genome Resource Grant AG 99-3205-8534  
 to H. A. Lewin and J. E. Mowack. Base Calling/Quality Scores: PHRED  
 from Washington University Genome Center. Vector Trimmi 97.  
 Cross-match from Washington University Genome Center PHRAP suite.  
 Sequences submitted are vector free and at least 200 bp in length.  
 PCR Primers:  
 FORWARD: TAAAGACCTTCACCTAAGG  
 BACKWARD: ATTACCTTCACCTAAG  
 Insert Length: 489 Std Error: 0.00  
 Plate: BP230016A10 row: E column: 9  
 Seq primer: AGCGATACCAATTTCACACGA  
 High quality sequence stop: 489.  
 Location/Qualifiers  
 1..489  
 /organism="Bos taurus"  
 /db\_xref="taxon:9913"  
 /clone\_lib="Soares normalized bovine placenta"  
 /sex="female"  
 /lab\_host="DH10B"  
 /note="Organ: placenta; Vector: pT73pac; Site:1; EcolR;  
 Site:2; NotI; The cDNA library was contributed by the  
 Soares laboratory and it was constructed and normalized  
 as described by Bonaldo, M.F., Lennon, G. and Soares,  
 M.B. (1996), Genome Research 6(9): 791-806. "  
 BASE COUNT 148 a 98 c 121 g 122 t  
 ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 489;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTCAGAAAGA 131  
 |||||||  
 DB 403 TAAAGACCTTCAGAAAGA 423

RESULT 10  
 BE198540 492 bp mRNA linear EST 26-JUN-2000  
 LOCUS u979905.y1 Soares\_mammary\_gland\_NMLMG Mus musculus cDNA clone  
 DEFINITION IMAGE:1548632 5', mRNA sequence.  
 ACCESSION BE198540  
 VERSION BE198540.1 GI:8710709  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclerognathii; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 492)  
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapsb-remail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:953980  
 Seq primer: -40RP from G4bco  
 High quality sequence stop: 475.  
 Location/Qualifiers  
 1..492  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_image="IMAGE:1548632"  
 /clone\_lib="Soares\_mammary\_gland\_NMLMG"  
 /sex="female (lactating)"  
 /tissue\_type="mammary gland"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from mammary  
 gland tissue from a lactating female, and was then primed  
 with a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptor (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pT73 vector. Library is normalized. Library  
 was constructed by Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 126 a 104 c 124 g 136 t  
 ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 492;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTCAGAAAGA 131  
 |||||||  
 DB 85 TAAAGACCTTCAGAAAGA 105

RESULT 11  
 AM579514 515 bp mRNA linear EST 16-MAR-2000  
 LOCUS MR0-H70164-070100-013-F05 HT0164 Homo sapiens CDNA, mRNA sequence.  
 DEFINITION AM579514  
 ACCESSION AM579514  
 VERSION AM579514.1 GI:7254563  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE HCCP http://www.ludwig.org.br/ORESTES.  
 AUTHORS The FAPESP/INIC Human Cancer Genome Project  
 TITLE The FAPESP/INIC Human Cancer Genome Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR0&t2=MR0-HT0164-070100-013-f05&t3=2000-01-07&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 7  
 High quality sequence stop: 515.  
 Location/Qualifiers

## FEATURES

source

```
1. .515
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0164"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT      157 a      103 c      124 g      131 t
ORIGIN
```

Query Match 13.7%; Score 21; DB 9; Length 515;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131  
 ||||||||||||||||||||  
 Db 393 TAAAGACCTTGCAAGAAAGA 413

RESULT 12  
 BM228157 560 bp mRNA linear EST 31-JAN-2002  
 LOCUS K0256B07-3 NIA Mouse Unfertilized Egg cDNA Library (Long) Mus  
 DEFINITION Musculus cDNA clone K0256B07.3, mRNA sequence.  
 ACCESSION BM228157  
 VERSION BM228157.1 GI:17790695  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 560)  
 Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Stagg, C.A., Martin, P., Alpa, K., Tanaka, T. and Ko, M.S.H.  
 Systematic Analyses of NIA Mouse Unfertilized Egg cDNA Library (Long)  
 Unpublished (2001)  
 JOURNAL Contact: Dawood B. Dudekula  
 COMMENT Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@igsun.grc.nia.nih.gov  
 Plate: K0256 row: B column: 07  
 Seq primer: -21M13 Forward  
 High quality sequence stop: 560  
 POLY-A=yes.

## FEATURES

source

```
1. .560
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="nia:EST:K0256B07-3"
/db_xref="taxon:10090"
/clone="K0256B07"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long)
```

```
)"
/tissue_type="Unfertilized Egg"
/lab_host="DH10B"
/note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544191]). Total RNAs were extracted from a pool of 1488 unfertilized eggs. Double-stranded cDNAs were synthesized with an oligo(dt) primer (Invitrogen): 5'-gacagatgcttgatgcgacgagccgcccttttttttttt-3', treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao (NIA)."
```

BASE COUNT 173 a 132 c 115 g 140 t  
 ORIGIN

Query Match 13.7%; Score 21; DB 10; Length 560;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 TAAAGACCTTGCAAGAAAGA 131  
 ||||||||||||||||||||  
 Db 480 TAAAGACCTTGCAAGAAAGA 460

RESULT 13  
 A1412929/C 589 bp mRNA linear EST 09-FEB-1999  
 LOCUS EST241229 Normalized rat kidney, Bento Soares Rattus sp. cDNA clone  
 DEFINITION R18EG55.3' end, mRNA sequence.  
 ACCESSION A1412929  
 VERSION A1412929.1 GI:4256433  
 KEYWORDS EST.  
 SOURCE Rattus sp.  
 ORGANISM Rattus sp.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 589)  
 Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J., Kerlavage, A.R. and Adams, M.D.  
 Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat Gene Index  
 Unpublished (1998)  
 JOURNAL Contact: Lee, NH  
 COMMENT The Institute for Genomic Research  
 9712, Medical Center Drive, Rockville, MD 20850, USA  
 Tel: (301)-838-3529  
 Fax: (301)-838-0208  
 Email: nhlee@igr.org  
 Seq primer: M13-21.

## FEATURES

source

```
1. .589
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="R18EG55"
/clone_lib="Normalized rat kidney, Bento Soares"
/note="Organ: kidney; Vector: pVT3pac; Site_1: EcoRI; Site_2: NotI"
```

BASE COUNT 200 a 140 c 105 g 144 t  
ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 589;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 111 TAAAGACCTTGCAAGAAAAGA 131  
|||||  
Db 497 TAAAGACCTTGCAAGAAAAGA 477

RESULT 14  
AM993220 594 bp mRNA linear EST 05-JUN-2000  
LOCUS RC2-BN0033-180200-014-f04 BN0033 Homo sapiens cDNA, mRNA sequence.  
DEFINITION AM993220  
VERSION AM993220.1 GI:8253368  
KEYWORDS EST  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 594)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W.Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=6t2-RC2-BN0033-180  
200-014-f04&t3=2000-02-18&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 20  
High quality sequence stop: 593.  
Location/Qualifiers  
1..594  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0033"  
/dev\_stage="Adult"  
/note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 179 a 116 c 135 g 164 t  
ORIGIN

Query Match 13.7%; Score 21; DB 9; Length 594;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 111 TAAAGACCTTGCAAGAAAAGA 131  
|||||

Db 417 TAAAGACCTTGCAAGAAAAGA 437

RESULT 15  
BF169086 597 bp mRNA linear EST 30-OCT-2000  
LOCUS 60177336AF1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3994215 5',  
DEFINITION mRNA sequence.  
ACCESSION BF169086  
VERSION BF169086.1 GI:11049438  
KEYWORDS EST  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 597)  
NIH-MGC http://mhc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: egadp@remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LML at:  
http://image.llnl.gov  
Plate: LLM9211 row: 0 column: 16  
High quality sequence stop: 585.  
Location/Qualifiers  
1..597  
/organism="Mus musculus"  
/strain="C57BL/6J (f1er1)"  
/db\_xref="taxon:10090"  
/clone\_lib="IMAGE:3994215"  
/clone\_lib="NCI\_CGAP\_Lu29"  
/tissue\_type="spontaneous tumor, metastatic to mammary.  
stem cell origin."  
/lab\_host="DH10P"  
/note="Organ: lung; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

FEATURES  
Source  
1..597  
Location/Qualifiers

BASE COUNT 175 a 106 c 143 g 173 t  
ORIGIN

Query Match 13.7%; Score 21; DB 10; Length 597;  
Best Local Similarity 100.0%; Pred. No. 1.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 111 TAAAGACCTTGCAAGAAAAGA 131  
|||||  
Db 45 TAAAGACCTTGCAAGAAAAGA 65

Search completed: August 2, 2002, 22:41:40  
Job time: 30155 sec

---







```
alignment_scores:
  Quality: 32.00      Length: 32
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-08-973-363-6 x AF006513

Align seg 1/1 to: AF006513 from: 1 to: 5947.

1 11leuProaspPaspProaspPlyslsProglAlaIalysGlnLeuGlnh 17
|||||
4025 ATTCTTCAGATGATCCGATATAAAACCAACAGAAACAGTTGCAGAC 4074
17 rArgAlaAspTyrLeuIleLysLeuSerArgAspLeuAlaLys 32
|||||
4075 CCGTGCAGACTACCTCATCAAAATTACTTAGTAGAGATCTTGCAAAA 4120

seq_name: gb-pr:AC092372

seq_documentation_block:
  LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001
  DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.
  ACCESSION AC092372
  VERSION AC092372.3 GI:17402768
  KEYWORDS htc.
  SOURCE human.
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  REFERENCE 1 (bases 1 to 101220)
  TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
  AUTHORS
  JOURNAL
  TITLE Direct Submission
  REFERENCE 2 (bases 1 to 101220)
  TITLE DOE Joint Genome Institute.
  AUTHORS
  JOURNAL
  TITLE Direct Submission
  REFERENCE 3 (bases 1 to 101220)
  TITLE Production Sequencing Facility, DOE Joint
  JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  REFERENCE 4 (bases 1 to 101220)
  TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
  JOURNAL Direct Submission
  COMMENT Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
  Drive, Walnut Creek, CA 94598, USA
  On Dec 7, 2001 this sequence version replaced gi:15290448.
  Draft Sequence Produced by DOE Joint Genome Institute
  WWW: jgi.doe.gov
  Finishing Completed at Stanford Human Genome Center
  WWW: shgc.stanford.edu
  Quality: Phred Quality >=40 100% of Sequence;
  Estimated Total Number of Errors is 0.
  NOTE: This insert is not the entire sequence of the clone (entire
  sequence is 146.7kb). It is clipped at the overlap with AC012624.
  The number of bases overlapped is 90404.
  Location/Qualifiers
    1..101220
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="RP11-58M12"
      /chromosome="5"

BASE COUNT 34122 a 18862 c 17827 g 30409 t
ORIGIN

alignment_scores:
  Quality: 32.00      Length: 32
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-08-973-363-6 x AC092372/rev

Align seg 1/1 to: reverse of: AC092372 from: 1 to: 101220

1 11leuProaspPaspProaspPlyslsProglAlaIalysGlnLeuGlnh 17
|||||
26932 ATTCTTCAGATGATCCGATATAAAACCAACAGAAACAGTTGCAGAC 26883
17 rArgAlaAspTyrLeuIleLysLeuSerArgAspLeuAlaLys 32
|||||
26882 CCGTGCAGACTACCTCATCAAAATTACTTAGTAGAGATCTTGCAAAA 26837

seq_name: gb-pr:AC012624

seq_documentation_block:
  LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
  DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
  ACCESSION AC012624
  VERSION AC012624.6 GI:14993679
  KEYWORDS htc.
  SOURCE human.
  ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
  REFERENCE 1 (bases 1 to 134365)
  TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
  AUTHORS
  JOURNAL
  TITLE Direct Submission
  REFERENCE 2 (bases 1 to 134365)
  TITLE DOE Joint Genome Institute.
  AUTHORS
  JOURNAL
  TITLE Direct Submission
  REFERENCE 3 (bases 1 to 134365)
  TITLE Production Sequencing Facility, DOE Joint
  JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
  REFERENCE 4 (bases 1 to 134365)
  TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
  JOURNAL Direct Submission
  COMMENT Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
  Drive, Walnut Creek, CA 94598, USA
  On Jul 21, 2001 this sequence version replaced gi:14277267.
  Location/Qualifiers
    1..134365
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="CTD-2082117"
      /chromosome="5"

BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN

alignment_scores:
  Quality: 32.00      Length: 32
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-08-973-363-6 x AC012624

Align seg 1/1 to: AC012624 from: 1 to: 134365

1 11leuProaspPaspProaspPlyslsProglAlaIalysGlnLeuGlnh 17
|||||
118247 ATTCTTCAGATGATCCGATATAAAACCAACAGAAACAGTTGCAGAC 118296
17 rArgAlaAspTyrLeuIleLysLeuSerArgAspLeuAlaLys 32
|||||
118297 CCGTGCAGACTACCTCATCAAAATTACTTAGTAGAGATCTTGCAAAA 118342

seq_name: gb_htg:AC021449

seq_documentation_block:
```

LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000  
 DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered  
 pieces.  
 AC021449  
 AC021449.3 GI:10047806  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 143079)  
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 Homo sapiens, clone RP11-58M12  
 Unpublished  
 2 (bases 1 to 143079)  
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
 Chepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,  
 Deatellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
 Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,  
 MacDonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K.,  
 McPeeters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,J.,  
 Norman,C.H., O'Connor,T., O'Donnell,P., Olyar,T.M., Peterson,K.,  
 Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,T.M., Rothman,D.,  
 Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,  
 Stojanovic,N., Sudramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
 Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
 Zimmer,A. and Zody,M.  
 Direct Submission  
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Sep 10, 2000 this sequence version replaced gi:1407963.  
 All repeats were identified using RepeatMasker:  
 Smit,A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html  
 ----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Project name: L5154  
 Center project name: 58\_M12  
 ----- Summary Statistics  
 Sequencing vector: M13; M77815; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 134743 bases at least Q40  
 Consensus quality: 139227 bases at least Q30  
 Consensus quality: 140814 bases at least Q20  
 Insert size: 144000; agarose-1p  
 Insert size: 142179; sum-of-contigs  
 Quality coverage: 4.6 in Q20 bases; agarose-1p  
 Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 38820: contig of 38820 bp in length  
 \* 38921 38920: gap of 100 bp  
 \* 38921 40411: contig of 1491 bp in length  
 \* 40412 40511: gap of 100 bp  
 \* 40512 43279: contig of 2768 bp in length  
 \* 43280 43379: gap of 100 bp

\* 43380 46905: contig of 3526 bp in length  
 \* 46906 47005: gap of 100 bp  
 \* 47006 51830: contig of 4825 bp in length  
 \* 51831 51930: gap of 100 bp  
 \* 51931 62619: contig of 10689 bp in length  
 \* 62620 62719: gap of 100 bp  
 \* 62720 75408: contig of 12689 bp in length  
 \* 75409 75508: gap of 100 bp  
 \* 75509 92516: contig of 17008 bp in length  
 \* 92517 92616: gap of 100 bp  
 \* 92617 106409: contig of 13793 bp in length  
 \* 106410 106509: gap of 100 bp  
 \* 106510 143079: contig of 36570 bp in length.  
 FEATURES  
 source  
 1. 143079  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-58M12"  
 /clone\_1b="RP11-58M12 Human Male BAC"  
 1. 38820  
 /note="assembly\_fragment"  
 /clone\_end:SP6  
 /vector\_side:left"  
 38921..40411  
 /note="assembly\_fragment"  
 40512..43279  
 /note="assembly\_fragment"  
 43380..46905  
 /note="assembly\_fragment"  
 47006..51830  
 /note="assembly\_fragment"  
 51931..62619  
 /note="assembly\_fragment"  
 62720..75408  
 /note="assembly\_fragment"  
 75509..92516  
 /note="assembly\_fragment"  
 92617..106409  
 /note="assembly\_fragment"  
 106510..143079  
 /note="assembly\_fragment"  
 /clone\_end:T7  
 /vector\_side:right"  
 BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
 ORIGIN  
 alignment\_scores:  
 Quality: 32.00 Length: 32  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-6 x AC021449 ..  
 Align seg 1/1 to: AC021449 from: 1 to: 143079  
 1 ILEuPProaSPaSPProaSPlySLySProGInAlaLySLnLaGInTh 17  
 116147 ATTCTTCAGATGATCCCGATPAAAAACCCACACAAACAGTTGCACAC 116196  
 17 TARGAlaSPtYrLeuAlleLySLeuleuSerArgASpLeuAlaLyS 32  
 116197 CCGGCGACAGACTACTCATCAATTAAGTAGAGAGATCTTCANAA 116242  
 seq\_name: gb\_htg:AC008531  
 seq\_documentation\_block:  
 LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
 7 ordered pieces.  
 AC008531  
 AC008531.3 GI:12830078

```

KEYWORDS      HTGS_PHASE2; HTGS_DRAFT.
SOURCE        humn.
ORGANISM      Homo sapiens
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE         DOE Joint Genome Institute.
REFERENCE     Sequencing of Human Chromosome 5
AUTHORS      Unpublished
TITLE         2 (bases 1 to 145659)
REFERENCE     DOE Joint Genome Institute.
AUTHORS      Direct Submission
TITLE         Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
JOURNAL       Genome Institute, 2800 Mitchell Drive Walnut Creek, CA 94598, USA
COMMENT       On Feb 14, 2001 this sequence version replaced gi:7528542.
-----Genome Center
Center Code: JGI
Web Site: http://www.jgi.doe.gov

Project Information
Center Project Name: 369535
Center Clone name: CIR-HSPC_480B11
-----
Summary Statistics
Consensus quality: 139128 bases at least Q40
Consensus quality: 142556 bases at least Q30
Consensus quality: 143744 bases at least Q20
Estimated insert size: 148000; pulse field gel estimation
Estimated insert size: 145059; sum-of-contigs estimation
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N's. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
56174: gap of 56174 bp in length
56175: gap of unknown length
56275: 100874: contig of 44600 bp in length
100875: 100974: gap of unknown length
100975: 113127: contig of 12153 bp in length
113128: 113127: gap of unknown length
113128: 118190: contig of 4963 bp in length
118191: 118290: gap of unknown length
118291: 119694: contig of 1404 bp in length
119695: 119794: gap of unknown length
119795: 123197: contig of 3503 bp in length
123198: 123197: gap of unknown length
123398: 145659: contig of 22262 bp in length.
*
Location/Qualifiers
1..145658
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CIRC-480B11"
/clone_lib="Caltech human BAC library C"
BASE COUNT   42561 a 26501 g 48609 t 600 cyr c*
ORIGIN
alignment_scores:
Quality:      32.00          Length:      32
Ratio:        1.000         Gaps:      0
Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x AC008531

```

```

Align seg 1/1 to: AC008531 from: 1 to: 145659

1 ||leleuProaspaspProkaspLyLysProGlnAlaLysGlnLeuGlnTh 17
|||||||||||||||||||||||||||||||||||||||||||||||||
46770 ATTCTTCGAGATGATCCGATTAATAAACCCAGCAAAACAGCTGGACAC 46819

17 rArgAlaAspTyrLeuIleLysIleuLeuSerArgAspLeuAlaLys 32
|||||||||||||||||||||||||||||||||||||||||||||||||
46820 CCGTCGACAGCTACCTCATCAATTAATTCTTGTAGATCTTGGCAAA 46865

seg_name: gb_htg.AC091946

seq_documentation block:
LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, *** SEQUENCING IN
PROGRESS ***, 33 unordered pieces.
ACCESSION AC091946
VERSION AC091946.1 GI:14333882
KEYWORDS HTG; HTGS; PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 193446)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 193446)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 544799
Center Clone name: RP01-11_36012
-----
Summary Statistics
Consensus quality: 157567 bases at least Q40
Consensus quality: 180259 bases at least Q30
Consensus quality: 184175 bases at least Q20
Estimated insert size: 204590; agarose-fp estimation
Estimated insert size: 190246; sum-of-contigs estimation
Quality coverage: 5.38 in Q20 bases; agarose-fp estimation
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1358: contig of 1358 bp in length
* 1359 1458: gap of unknown length
* 1459 2667: contig of 1209 bp in length
* 2668 2768: gap of unknown length
* 2768 3823: contig of 1056 bp in length
* 3823 3923: gap of unknown length
* 3923 5920: contig of 1997 bp in length
* 5920 6020: gap of unknown length
* 6021 7434: gap of unknown length
* 7435 7536: gap of unknown length
* 7536 9096: contig of 1502 bp in length
* 9097 9197: gap of unknown length
* 9197 10260: contig of 1064 bp in length
* 10261 10360: gap of unknown length
* 10361 12460: contig of 2100 bp in length
* 12461 12560: gap of unknown length

```



LOCUS AC022121 219258 bp DNA linear PRI 30-AUG-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.  
 ACCESSION AC022121  
 VERSION AC022121.6 GI:15375145  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 219258)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 219258)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 REFERENCE 3 (bases 1 to 219258)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 REFERENCE 4 (bases 1 to 219258)  
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 COMMENT On Aug 30, 2001 this sequence version replaced gi:15148108.  
 Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.8% of Sequence;  
 Estimated Total Number of Errors is 0.4.  
 STS Content:  
 WI-5811.G04974  
 WI-13675.G23101  
 SHGC-58345.G38487  
 SHGC-103595.G57841.  
 FEATURES  
 location/Qualifiers  
 SOURCE 1 219258  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2007H13"  
 BASE COUNT 71954 a 42062 c 40933 g 64309 t  
 ORIGIN  
 alignment\_scores:  
 Quality: 32.00 Length: 32  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 us-08-973-363-6 x AC022121/rev ..  
 Align seg 1/1 to reverse of: AC022121 from: 1 to: 219258  
 1 TleauPProbaSPProbaSPlystysProGlnAlaLysGlnleangnph 17  
 ||||||||||||||||||||||||||||||||||||||||||||  
 89450 ATTCTTCAGATGATTCGATGATGATGATGATGATGATGATGAT 89401  
 ||||||||||||||||||||||||||||||||||||||||||||  
 17 TArgAlaSPyTleuTleuTleuTleuTleuTleuTleuTleuTleuTleu 32  
 ||||||||||||||||||||||||||||||||||||||||||||  
 89400 CCGTCAGACTACTCTCATTAATTCTTAGTAAGATCTTGCAAA 89355  
 ||||||||||||||||||||||||||||||||||||||||||||  
 seq\_name: gb\_htg:AC092382  
 seq\_documentation\_block:  
 LOCUS AC092382 276181 bp DNA linear HTG 03-JUL-2001  
 DEFINITION Homo sapiens chromosome 5 clone RP11-75H1, WORKING DRAFT SEQUENCE,

47 unordered pieces.  
 AC092382  
 AC092382.1 GI:14589571  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 276181)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Human Chromosome 5  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 276181)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov  
 -----  
 Project Information  
 Center Project Name: 435334  
 Center clone name: RPT-11\_75H1  
 -----  
 Summary Statistics  
 Consensus quality: 229677 bases at least Q40  
 Consensus quality: 256163 bases at least Q30  
 Consensus quality: 260799 bases at least Q20  
 Estimated insert size: 174820; agarose-fp estimation  
 Estimated insert size: 271581; sum-of-contigs estimation  
 Quality coverage: 8.62 in Q20 bases; agarose-fp estimation  
 Quality coverage: 5.55 in Q20 bases; sum-of-contigs estimation.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 47 contigs, the true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 1061: contig of 1061 bp in length  
 \* 1062 1161: gap of unknown length  
 \* 1162 2827: contig of 1666 bp in length  
 \* 2828 2927: gap of unknown length  
 \* 2928 4227: contig of 1300 bp in length  
 \* 4228 4327: gap of unknown length  
 \* 4328 5618: contig of 1291 bp in length  
 \* 5619 5718: gap of unknown length  
 \* 5719 6983: contig of 1265 bp in length  
 \* 6984 7083: gap of unknown length  
 \* 7084 8422: contig of 1339 bp in length  
 \* 8423 8522: gap of unknown length  
 \* 8523 9771: contig of 1249 bp in length  
 \* 9772 9871: gap of unknown length  
 \* 9872 11198: contig of 1327 bp in length  
 \* 11199 11298: gap of unknown length  
 \* 11299 12414: contig of 1116 bp in length  
 \* 12415 12514: gap of unknown length  
 \* 12515 13806: contig of 1292 bp in length  
 \* 13807 13906: gap of unknown length  
 \* 13907 15360: contig of 1454 bp in length  
 \* 15361 15460: gap of unknown length  
 \* 15461 16667: contig of 1207 bp in length  
 \* 16668 16767: gap of unknown length  
 \* 16768 18390: contig of 1623 bp in length  
 \* 18391 18491: gap of unknown length  
 \* 18492 19641: contig of 1051 bp in length  
 \* 19642 21156: contig of 1515 bp in length  
 \* 21157 21256: gap of unknown length  
 \* 21257 22448: contig of 1192 bp in length

```
* 22449 22548: gap of unknown length
* 22549 22499: contig of 1951 bp in length
* 24500 24599: gap of unknown length
* 24600 26384: contig of 1785 bp in length
* 26385 26484: gap of unknown length
* 26485 27883: contig of 1399 bp in length
* 27884 27983: gap of unknown length
* 27984 30386: contig of 2403 bp in length
* 30387 30486: gap of unknown length
* 30487 32480: contig of 1994 bp in length
* 32481 32580: gap of unknown length
* 32581 35441: contig of 2861 bp in length
* 35442 35541: gap of unknown length
* 35542 37800: contig of 2259 bp in length
* 37801 37900: gap of unknown length
* 37901 40587: contig of 2687 bp in length
* 40588 40687: gap of unknown length
* 40688 44159: contig of 3472 bp in length
* 44160 44259: gap of unknown length
* 44260 46336: contig of 2377 bp in length
* 46337 46736: gap of unknown length
* 46737 50082: contig of 3346 bp in length
* 50083 50182: gap of unknown length
* 50183 53988: contig of 3806 bp in length
* 53989 54088: gap of unknown length
* 54089 56592: contig of 2504 bp in length
* 56593 61352: gap of unknown length
* 61353 61452: gap of unknown length
* 61453 64254: contig of 2802 bp in length
* 64255 64354: gap of unknown length
* 64355 68825: contig of 4471 bp in length
* 68826 68925: gap of unknown length
* 68926 74393: contig of 5468 bp in length
* 74394 74493: gap of unknown length
* 74494 80561: contig of 6068 bp in length
* 80562 80661: gap of unknown length
* 80662 87626: contig of 6965 bp in length
* 87627 87726: gap of unknown length
* 87727 93600: contig of 5874 bp in length
* 93601 93700: gap of unknown length
* 93701 102024: contig of 8324 bp in length
* 102025 102124: gap of unknown length
* 102125 109905: contig of 7781 bp in length
* 109906 110005: gap of unknown length
* 110006 116737: contig of 6732 bp in length
* 116738 116837: gap of unknown length
* 116838 126797: contig of 9960 bp in length
* 126798 126897: gap of unknown length
* 126898 135114: contig of 8217 bp in length
* 135115 135214: gap of unknown length
* 135215 146713: contig of 11499 bp in length
* 146714 146813: gap of unknown length
* 146814 157894: contig of 11081 bp in length
* 157895 157994: gap of unknown length
* 157995 171752: contig of 13758 bp in length
* 171753 171852: gap of unknown length
* 171853 201687: contig of 29835 bp in length
* 201688 201787: gap of unknown length
* 201788 231268: contig of 29481 bp in length
* 231269 231368: gap of unknown length
* 231369 276181: contig of 44813 bp in length.
```

FEATURES  
source

```
1..276181
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-75H1"
/clone.lib="RP1 human BAC library 11"
BASE COUNT 86859 a 51769 c 50026 g 82854 t 4673 others
ORIGIN
```

```
alignment_scores:
      quality: 32.00      length: 32
      ratio: 1.000      gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
```

## alignment\_block:

US-08-973-363-6 x AC092382/rev ..

Align seg 1/1 to reverse of: AC092382 from: 1 to: 276181

```
1 1lleuProaspaspProaspLysProglAlaLysGlnLeuGln 17
|||||
213694 ATTCTCCAGATGATCCGATFAAAAAACCAACCAACAGATTGCAGAC 213645
```

```
17 rArgAlaAspTyrIleuIleLysLeuLeuSerArgAspLeuAlaLys 32
|||||
213644 CCGTGCAGACTACCTCATTAATTACTTAGTAGAGATTGCAGAAA 213599
```

seq\_name: gb\_pat:A58696

seq\_documentation\_block:

LOCUS A58696 1316 bp DNA linear PAT 06-MAR-1998

DEFINITION Sequence 15 from Patent W09639505.

ACCESSION A58696

VERSION A58696.1 GI:3714253

KEYWORDS

SOURCE

ORGANISM

unidentified.

unclassified.

REFERENCE

1 (bases 1 to 1316)

AUTHORS

Griffiths, R. and Tiwari, B.

TITLE

AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS

JOURNAL

Patent: WO 9639505-A 15 12-DEC-1996;

ISIS INNOVATION (GB)

Other publication AU 5906996 961224.

COMMENT

Location/Qualifiers

FEATURES

1..1316

/organism="unidentified"

/db\_xref="taxon:32644"

BASE COUNT 493 a 205 c 308 g 304 t 6 others

ORIGIN

alignment\_scores:

Quality: 26.00 length: 26

Ratio: 1.000 gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x A58696 ..

Align seg 1/1 to: A58696 from: 1 to: 1316

```
1 1lleuProaspaspProaspLysProglAlaLysGlnLeuGln 17
|||||
1138 ATTTCCTGATGATCCAGATAGAAACCCAGGCTAACGATTACAGAC 1187
```

```
17 rArgAlaAspTyrIleuIleLysLeuLeuSerArgAspLeuAlaLys 26
|||||
1188 TCGTGCAGATTACCTCATTAATTACTG 1215
```

seq\_name: gb\_pat:A58691

seq\_documentation\_block:

LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998

DEFINITION Sequence 10 from Patent W09639505.

ACCESSION A58691

VERSION A58691.1 GI:3714250

KEYWORDS

SOURCE

unidentified.

unclassified.

ORGANISM

unclassified.

unclassified.



```

source 1.1311
BASE COUNT 441 a 257 c 316 g 297 t
ORIGIN

alignment_scores:
  Quality: 25.00      Length: 25
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x AR029026 ..

Align seg 1/1 to: AR029026 from: 1 to: 1311

1 11leuProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17
|||||
1237 ATTCTTCCAGATGATCCTGATATAAACCACAAACAAACAGTTACGAC 1286
17 rArgAlaaspTYrLeu11leuLysLeu 25
|||||
1287 CCCTGCAGACTACCTCATCAACTA 1311

```





8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuLeu 24  
 |||  
 52 AAAAAACCAAGCAACAGTTACAGACCCGTCAGACTACTCTATCAA 101  
 |||  
 24 sLeuLeuSerArgAspLeuAlaLysArgGluAlaGlnArgLeuCysGlyA 41  
 |||  
 102 ACTACTTACGACAGATCTTGCAGAAAAGAGAGCGCTCAGAGACTTTGTGTG 151  
 |||  
 41 la 41  
 ||  
 152 CG 153

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42754

seq\_documentation\_block:

ID AAAT42754 standard; cDNA: 1316 BP.

AC AAAT42754;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-W gene (partial sequence).

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

PN WO9639505-A1.

PD 12-DEC-1996.

PE 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI: 1997-043127/04.

PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny

PS Claim 1; Fig 8; 76pp: English.

CC The chicken CHD-W gene (AAAT42754) acting alone or in conjunction with  
 CC the closely related CHD-1A gene (AAAT42751) is suggested to initiate  
 CC female development in birds. The sequence of CHD-W was deduced  
 CC from 2 clones isolated from a 10-day chick embryo library using  
 CC a fragment of the CHD-1A gene as a probe. The CHD-W gene is  
 CC located on the W chromosome. Probes based on CHD-W and CHD-1A give  
 CC a W chromosome-specific signal on hybridisation to genomic DNA of a  
 CC non-rafite bird and can be used for sex determin. of a bird. CHD-W  
 CC nucleic acids can also be used to control the sex of progeny of a  
 CC bird.

SO Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

alignment\_scores:

Quality: 26.00 Length: 26  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AAAT42754 ..

Align seg 1/1 to: AAAT42754 from: 1 to: 1316

1 lIleuProAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17

|||||  
 1138 ATTTCCTGATGATCCAGATTAAGAAACCCAGGCTACAGTTACAGAC 1187  
 17 fArgAlaAspTyrLeuIleLysLeu 26  
 |||  
 1188 TCCTGCGATTAATCTTAAATTAATCTG 1215

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAAT42751

seq\_documentation\_block:

ID AAAT42751 standard; cDNA: 6608 BP.

AC AAAT42751;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-1A gene.

KW Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

FN Key Location/Qualifiers

FT CDS 228..5390

/\*tag= a

PN WO9639505-A1.

PD 12-DEC-1996.

PE 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI: 1997-043127/04.

PT Avian chromodomain-helicase-DNA binding genes determine sex in  
 PT birds - used for sex determ. and to control sex of progeny

PS Claim 1; Fig 5; 76pp: English.

CC The chicken CHD-W gene (AAAT42754) acting alone or in conjunction with  
 CC the closely related CHD-1A gene (AAAT42751) is suggested to initiate  
 CC female development in birds. The sequence of CHD-1A was deduced  
 CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
 CC library using a great tit CHD-W sequence (see also AAAT42755) as probe.  
 CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
 CC gene (see also AAAT42756-57). It is located on an autosome or X  
 CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
 CC specific signal on hybridisation to genomic DNA of a non-rafite  
 CC bird and can be used for sex determin. of a bird. CHD-1A nucleic  
 CC acids can also be used to control the sex of the progeny of a bird.

SO Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores:

Quality: 26.00 Length: 26  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AAAT42751 ..

Align seg 1/1 to: AAAT42751 from: 1 to: 6608

1 lIleuProAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
 |||

4080 ATTTCACCTGATGATCCAGACAGAAACCCAGGCAAGACGCTACAGAC 4129

17 TARGALASPTRYLEUILELYSLEU 26

4130 CCGTCGACGACTACCTCATTTAAATTCATG 4157

seq\_name: /SIDS1/gcgdata/hold-geneseg/genesegn-emb1/NA1998.DAT:AAV59280

seq\_documentation\_block:

ID AAV59280 standard; CDNA: 1311 BP.

AC AAV59280;

DT 14-DEC-1998 (first entry)

DE Altered telomere repeat binding factor 1 gene.

DS ds: human; telomere repeat binding factor; A-TRF; dimerisation domain;

KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

OS Homo sapiens.

OS Synthetic.

FT CDS

FT 1,1311

FT /product- "A-TRF"

PN MO9836066-A1.

PD 20-AUG-1998.

PF 13-FEB-1998; 98MO-US02765.

PR 04-FEB-1998; 9805-0018628.

PR 13-FEB-1997; 9705-0800264.

PA (UYRQ) UNIV ROCKEFELLER.

PI Bianchi A, De Lange T, Van Steensel B;

DR WPI: 1998-480769/41.

DR P-PSDB: AAM59280.

PT Nucleic acid encoding altered telomere repeat binding protein and

PT related vectors - transformants, hetero-dimers and antibodies, used

PT to inhibit shortening of telomerases caused by ageing or disease,

PT also used to extend life of cells in culture

PS Claim 14: Page 110-111; 163pp; English.

XX The altered vertebrate telomere repeat binding protein (A-TRF) has a

XX telomere repeat binding factor (TRF) dimerisation domain, and forms a

XX hetero-dimer with TRF, preventing it from binding to the specified repeat

XX sequence. A-TRF, optionally expressed with gene therapy, is used to

XX inhibit shortening of telomeres associated with ageing (for cosmetic

XX purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,

XX atrophy of the skin, age-related macular degeneration, atherosclerosis,

XX tumours and viral (including human immune deficiency virus) infection.

XX Cells expressing A-TRF also have an increased life span in vitro, e.g.

XX for expression of recombinant proteins or where intended for subsequent

XX transplant or for testing, eliminating the need for transformation.

XX Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

XX

alignment\_scores:

Quality: 25.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AAV59280

Align seg 1/1 to: AAV59280 from: 1 to: 1311

1 TleleuproaspaspProaspIysIysProctlnAlaygInleuInth 17

1237 ATTCTTCGACGATGATCCGATTAATAAACCCAGCAAGCAAGACTTACAGAC 1286

17 TARGALASPTRYLEUILELYSLEU 25

1287 CCGTCGACGACTACCTCATTTAAATTCATG 1311

seq\_name: /SIDS1/gcgdata/hold-geneseg/genesegn-emb1/NA1997.DAT:AAAT42759

seq\_documentation\_block:

ID AAT42759 standard; DNA: 153 BP.

AC AAT42759;

DT 12-MAR-1997 (first entry)

DE Great tit CHD-W gene fragment.

DS Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KW CHD-1A; CHD-W; W chromosome; ss.

OS Parus major.

OS

FT Key

FT msc\_difference 52..81

FT /tag- a

FT /note- "bases 52-81 are a repeat of bases 22-51

FT and are ignored in the translated amino

FT acid sequence given in Fig 3"

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96MO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI: 1997-043127/04.

DR P-PSDB: AAM08149.

PT Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determ. and to control sex of progeny

PS Claim 8: Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology

XX to portions of the chicken CHD-1A (A - Avian) gene (AAT42757),

XX chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and

XX and the great tit CHD-W gene (AAT42759). Translated amino acid

XX sequences of this region are provided in AAM08146-49. The CHD-1A

XX (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine

XX sex in birds and can be used to identify the sex of an embryo,

XX foetus etc. and to manipulate the sex of progeny.

XX Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

XX

alignment\_scores:

Quality: 19.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AAT42759

Align seg 1/1 to: AAT42759 from: 1 to: 153

```
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleIle 24
|||||
52 AAGAAACCCAGGCAAGACAGTGTACAGACCGCTGACATTCATTTAA 101
24 sleuLeu 26
|||||
102 ATTACTG 108
```

seq\_name: /SIDSI/gcgdata/hold-geneseg/geneseqn-emb1/NA1997.DAT:AAT42757

seq\_documentation\_block:

ID AAT42757 standard; DNA; 153 BP.

XX AAT42757;

XX 12-MAR-1997 (first entry)

XX Chick CHD-1A gene fragment.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;

XX CHD-1A; CHD-W; W chromosome; ss.

XX Gallus sp.

XX Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag- a

FT /note- "bases 52-81 are a repeat of bases 22-51 and are ignored in the translated amino acid sequence given in Fig 3"

XX WO9639505-A1.

XX 12-DEC-1996.

XX 05-JUN-1996; 96WO-GB01341.

XX 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX Griffiths R, Tiwari B;

XX WPI; 1997-043127/04.

XX P-PSDB; AAW08147.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

XX birds - used for sex determ. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757).  
CC chicken CHD-W (W refers to the W chromosome) gene (see also AAT42758)  
CC and the great tit CHD-W gene (see also AAT42759). Translated amino  
CC acid sequences of this region are provided in AAW08146-49. The  
CC CHD-1A (see also AAT42751) and CHD-W (see also AAT42754-55) genes  
CC determine sex in birds and can be used to identify the sex of an  
CC embryo, foetus etc. and to manipulate the sex of progeny.

XX SQ Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment\_scores:

Quality: 19.00

Length: 19

Ratio: 1.000

Gaps: 0

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AAT42757

Align seg 1/1 to: AAT42757 from: 1 to: 153

```
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleIle 24
|||||
52 AAGAAACCCAGGCAAGACAGTGTACAGACCGCTGACATTCATTTAA 101
24 sleuLeu 26
|||||
102 ATTACTG 108
```

seq\_name: /SIDSI/gcgdata/hold-geneseg/geneseqn-emb1/NA1997.DAT:AAT42758

seq\_documentation\_block:

ID AAT42758 standard; DNA; 153 BP.

XX AAT42758;

XX 12-MAR-1997 (first entry)

XX Chick CHD-W gene fragment.

XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;

XX CHD-1A; CHD-W; W chromosome; ss.

XX Gallus sp.

XX Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag- a

FT /note- "bases 52-81 are a repeat of bases 22-51 and are ignored in the translated amino acid sequence given in Fig 3"

XX WO9639505-A1.

XX 12-DEC-1996.

XX 05-JUN-1996; 96WO-GB01341.

XX 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

XX Griffiths R, Tiwari B;

XX WPI; 1997-043127/04.

XX P-PSDB; AAW08148.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

XX birds - used for sex determ. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (AAT42757).  
CC chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
CC and the great tit CHD-W gene (AAT42759). Translated amino acid  
CC sequences of this region are provided in AAW08146-49. The CHD-1A  
CC (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
CC sex in birds and can be used to identify the sex of an embryo,  
CC foetus etc. and to manipulate the sex of progeny.

XX SQ Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment\_scores:

Quality: 19.00

Length: 19

Ratio: 1.000

Gaps: 0

Percent Similarity: 100.000

Percent Identity: 100.000

alignment\_block:

US-08-973-363-6 x AAT42758

```

Align seg 1/1 to: AAT42758 from: 1 to: 153
      8 LysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleTy 24
      11 |||||||
      52 AAGAAACCCGAGCTAAGCAGTTCAGACCCGTCGAGATTCATTAA 101
      24 sLeuLeu 26
      11 |||||||
      102 ATTACTG 108

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06443
seq_documentation_block:
ID   ABL06443 standard; cDNA; 6240 BP.
XX
AC   ABL06443;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide seq ID NO 13811.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB62340.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 13811; 21bp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABB57737-ABB72072).
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ   Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

alignment_scores:
      Quality: 11.00      Length: 11
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

      9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19
      11 |||||||
      4205 AAGCCCCAGCCGAGCAGCTGACGACGCGTCCC 4237
      seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06442
seq_documentation_block:
ID   ABL06442 standard; cDNA; 9933 BP.
XX
AC   ABL06442;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide seq ID NO 13808.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB62339.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 13808; 21bp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABB57737-ABB72072).
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ   Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;

alignment_scores:
      Quality: 11.00      Length: 11
      Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-973-363-6 x ABL06442 ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933

      9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19
      11 |||||||
      6745 AAGCCCCAGCCGAGCAGCTGACGACGCGTCCC 6777
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AKB88882

```

seq\_documentation\_block:  
ID AAK8882 standard; cDNA; 421 BP.  
AC AAK8882;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen coding sequence SEQ ID NO: 1198.  
XX  
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 20000US-0250391.  
 PR 05-DEC-2000; 20000US-0251030.  
 PR 05-DEC-2000; 20000US-0251868.  
 PR 05-DEC-2000; 20000US-0256719.  
 PR 06-DEC-2000; 20000US-0251479.  
 PR 08-DEC-2000; 20000US-0251856.  
 PR 08-DEC-2000; 20000US-0251869.  
 PR 08-DEC-2000; 20000US-0251868.  
 PR 08-DEC-2000; 20000US-0251989.  
 PR 08-DEC-2000; 20000US-0251990.  
 PR 11-DEC-2000; 20000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-502630/55.  
 DR P-PSDB; AAM93109.  
 PT Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognosing disorders of the  
 PT digestive system, particularly cancer and cancer metastases -  
 XX  
 PS Claim 1; SEQ ID NO 1198; 986pp; English.  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer, Meckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis. The present sequence is a cDNA encoding a digestive  
 CC system antigen of the invention.  
 SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
 XX  
 alignment\_scores:  
 Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-6 x AAK88882 ..  
 Align seq 1/1 to: AAK88882 from: 1 to: 421  
 13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
 ||||||||||||||||||||||||||||  
 146 AAGCAGCTACGAGACCGGACGATTAATTG 175  
 seq\_name: /SIISI1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AA157603  
 seq\_documentation\_block:  
 ID ID AA157603 standard; cDNA; 421 BP.  
 XX  
 AC AA157603;  
 XX  
 DT .19-OCT-2001 (first entry)  
 XX  
 DE Human colorectal cancer antigen cDNA SEQ ID NO: 67.  
 XX  
 KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.  
 OS Homo sapiens.  
 XX  
 PN W020015350-A1.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01350.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR

PR	14-FEB-2000	2000US	-0.186568
PR	24-FEB-2000	2000US	-0.184654
PR	02-MAR-2000	2000US	-0.163650
PR	15-MAR-2000	2000US	-0.169874
PR	17-MAR-2000	2000US	-0.190767
PR	18-MAR-2000	2000US	-0.198133
PR	19-MAY-2000	2000US	-0.205515
PR	07-JUN-2000	2000US	-0.209457
PR	28-JUN-2000	2000US	-0.211486
PR	30-JUN-2000	2000US	-0.215135
PR	07-JUL-2000	2000US	-0.215670
PR	11-JUL-2000	2000US	-0.216580
PR	11-JUL-2000	2000US	-0.217487
PR	14-JUL-2000	2000US	-0.218426
PR	26-JUL-2000	2000US	-0.220964
PR	14-AUG-2000	2000US	-0.224519
PR	14-AUG-2000	2000US	-0.225214
PR	14-AUG-2000	2000US	-0.225214
PR	14-AUG-2000	2000US	-0.225267
PR	14-AUG-2000	2000US	-0.225268
PR	14-AUG-2000	2000US	-0.225270
PR	14-AUG-2000	2000US	-0.225447
PR	14-AUG-2000	2000US	-0.225457
PR	14-AUG-2000	2000US	-0.225758
PR	14-AUG-2000	2000US	-0.226279
PR	18-AUG-2000	2000US	-0.226279
PR	22-AUG-2000	2000US	-0.226661
PR	22-AUG-2000	2000US	-0.226862
PR	22-AUG-2000	2000US	-0.227182
PR	23-AUG-2000	2000US	-0.228709
PR	30-AUG-2000	2000US	-0.228974
PR	01-SEP-2000	2000US	-0.229287
PR	01-SEP-2000	2000US	-0.229343
PR	01-SEP-2000	2000US	-0.229344
PR	01-SEP-2000	2000US	-0.229344
PR	05-SEP-2000	2000US	-0.229509
PR	05-SEP-2000	2000US	-0.229513
PR	06-SEP-2000	2000US	-0.230437
PR	06-SEP-2000	2000US	-0.230442
PR	08-SEP-2000	2000US	-0.231248
PR	08-SEP-2000	2000US	-0.231244
PR	08-SEP-2000	2000US	-0.231414
PR	08-SEP-2000	2000US	-0.231414
PR	08-SEP-2000	2000US	-0.232080
PR	08-SEP-2000	2000US	-0.232080
PR	12-SEP-2000	2000US	-0.231967
PR	14-SEP-2000	2000US	-0.231398
PR	14-SEP-2000	2000US	-0.231398
PR	14-SEP-2000	2000US	-0.231398
PR	14-SEP-2000	2000US	-0.233401
PR	14-SEP-2000	2000US	-0.233401
PR	14-SEP-2000	2000US	-0.233406
PR	14-SEP-2000	2000US	-0.233406
PR	14-SEP-2000	2000US	-0.233463
PR	21-SEP-2000	2000US	-0.234223
PR	21-SEP-2000	2000US	-0.234274
PR	25-SEP-2000	2000US	-0.234958
PR	25-SEP-2000	2000US	-0.234958
PR	27-SEP-2000	2000US	-0.235893
PR	27-SEP-2000	2000US	-0.235893
PR	29-SEP-2000	2000US	-0.235876
PR	29-SEP-2000	2000US	-0.235868
PR	29-SEP-2000	2000US	-0.236357
PR	29-SEP-2000	2000US	-0.236359
PR	02-OCT-2000	2000US	-0.237002
PR	02-OCT-2000	2000US	-0.237002

PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0246174.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0255678.  
XX  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-457727/49.  
XX  
XX P-PSDB; AAM38625.  
XX  
XX  
XX Isolated polypeptide for treating, preventing and/ or prognosing  
XX  
XX PT disorders related to the colon and rectum including colorectal cancers  
XX  
XX PT and also for testing and detection e.g. diagnosis -

XX  
XX Claim 1: SEQ ID NO: 67; 522pp + Sequence Listing: English.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX  
XX number of colorectal cancer antigens. These are shown in  
XX  
XX CC AA157547-AA157619 and AAM38569-AA38641. These can be used in the  
XX  
XX CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.  
XX  
XX CC The present sequence is a colorectal cancer antigen coding sequence of  
XX  
XX CC the invention.  
XX  
XX CC Note: The sequence data for this patent did not form part of the printed  
XX  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX  
XX CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX  
XX SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
  
alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment block:  
US-08-973-363-6 x AA157603 ..  
  
Align seq 1/1 to: AA157603 from: 1 to: 421  
  
13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
|||||  
146 AAGCAGCTACGACCCGAGCGATTACTTG 175  
  
seq\_name: /SIDSI/gcsgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: AAC99822  
  
seq\_documentation\_block:  
ID AAC99822 standard; cDNA; 1148 BP.  
XX  
XX AAC99822;  
XX  
XX 13-MAR-2001 (first entry)  
XX  
XX DE Human secreted protein gene 5 SEQ ID NO:15.  
XX  
XX  
XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
XX  
XX antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;  
XX  
XX cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;  
XX  
XX fungicide; ophthalmological; gene therapy; pathological condition;  
XX  
XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;  
XX  
XX neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;  
XX  
XX cerebrovascular disorder; angiogenesis; nervous system disorder;  
XX  
XX Alzheimer's disease; infection; ocular disorder; corneal infection;  
XX  
XX wound healing; skin aging; food additive; preservative; ss.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX  
XX WO200070042-A1.  
XX  
XX PD 23-NOV-2000.  
XX  
XX  
XX 11-MAY-2000; 2000WO-US12788.  
XX  
XX PF  
XX  
XX 13-MAY-1999; 99US-0134068.  
XX  
XX  
XX  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;  
XX  
XX PI Duan RD, Florence KA, Soppet DR;  
XX  
XX  
XX WPI; 2000-679828/66.  
XX  
XX DR P-PSDB; AAB56081.  
XX  
XX  
XX  
XX Isolated nucleic acid molecule encoding a human secreted protein is  
XX  
XX PT used in preventing, treating or ameliorating a medical condition -  
XX  
XX  
XX Claim 1; Page 858; 1065pp; English.



XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the  
CC human secreted proteins given in AAB56077 to AAB56562. Human secreted  
CC proteins have activities based on the tissues and cells the genes are  
CC expressed in. Examples of activities include: immunosuppressive;  
CC antitachytic; antithrombotic; antiproliferative; cytostatic; cardiant;  
CC vasotrophic; cerebroprotective; neurotrophic; neuroprotective; antibacterial;  
CC virucide; fungicide; and opthalmalological. The human secreted  
CC polynucleotides and proteins can be used to prevent, treat or ameliorate  
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,  
CC dogs, chickens or sheep. They are also used in diagnosing a pathological  
CC condition or susceptibility to a pathological condition. Disorders which  
CC are diagnosed or treated include autoimmune diseases e.g. Rheumatoid  
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders  
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
CC fungi and ocular disorders e.g. corneal infection. The proteins can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotherapy. The proteins can also be used as a  
CC food additive or preservative to increase or decrease storage  
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used  
CC in the exemplification of the present invention.  
XX  
SQ Sequence 1148 BP; 353 A; 251 C; 221 G; 322 T; 1 other:  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-6 x AAC99822 ..  
  
Align seg 1/1 to: AAC99822 from: 1 to: 1148  
  
27 SerArgAspLeuAlaIysArgIu 34  
|||||  
162 TCTAGGATCTACGCAAGCGAGG 185  
  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:AAS96223  
  
seq\_documentation\_block:  
ID AAS96223 standard; CDNA; 1973 BP.  
XX  
AC AAS96223;  
XX  
DF 26-FEB-2002 (first entry)  
XX  
DE Human bone marrow tissue polynucleotide #9.  
XX  
KW Human; bone marrow; central nervous system disorder; haematopoesis; ss;  
KW peripheral nervous system disorder; myeloid cell disorder; osteoporosis;  
KW lymphoid cell disorder; osteoarthritis; bone degenerative disorder; skin;  
KW periodontal disease; cerebrovascular disorder; viral infection; cancer;  
KW bacterial infection; fungal infection; autoimmune disorder; pancreas;  
KW hyperproliferative disorder; cognitive depression; stomach;  
KW inflammatory disorder; tumour; colon; liver; bladder; prostate; kidney;  
KW ovary; brain; skin; bone; cartilage; tendon; ligament; nerve;  
KW tissue growth; tissue regeneration; wound; burn; incision; ulcer.  
XX  
OS Homo sapiens.  
XX  
PN WO200179447-A2.  
XX  
PD 25-OCT-2001.  
XX  
PF 18-APR-2001; 2001WO-US12607.  
XX  
PR 18-APR-2000; 2000US-0522929.

PR 24-OCT-2000; 2000US-0695783.  
PR 30-NOV-2000; 2000US-250583P.  
PR 26-JAN-2001; 2001US-0770160.  
XX  
PA (HYSE-) HYSED INC.  
PI Ford JE, Boyle BJ, Tang YT, Asundi V, Yang Y, Liu C, Drmanac RT;  
XX WPI; 2001-626527/72.  
DR P-PSDB; AAU71931.  
XX  
PT Polynucleotides encoding bone marrow tissue polypeptides, useful for  
XX preventing, diagnosing and treating osteopathic diseases -  
XX  
PS Claim 1: Page 137-138; 202pp; English.  
XX  
XX The invention relates to isolated polynucleotides encoding bone marrow  
CC tissue polypeptides. The sequences of the invention can be used in the  
CC treatment of central nervous system disorders such as Parkinson's  
CC disease, Alzheimer's disease, Huntington's disorders and amyotrophic  
CC lateral sclerosis, peripheral nervous system disorders such as peripheral  
CC neuropathy, haematopoesis, myeloid and lymphoid cell disorders,  
CC osteoporosis, osteoarthritis, bone degenerative disorders, periodontal  
CC disease, cerebrovascular disorders such as stroke, viral infections,  
CC bacterial infections, fungal infections, autoimmune disorders such as  
CC rheumatoid arthritis, multiple sclerosis, asthma and eczema, inflammatory  
CC disorders such as Crohn's disease and nephritis, hyperproliferative  
CC disorders such as psoriasis, cognitive disorders, depression and cancer  
CC (e.g. tumours of the stomach, colon, pancreas, liver, bladder, prostate,  
CC ovary, kidney, brain and skin). The sequences are also useful in bone,  
CC cartilage, tendon, ligament and/or nerve tissue growth or regeneration  
CC and in healing of wounds, burns, incisions and ulcers. Sequences  
CC AAC96215-AAS96271 represent cDNA molecules encoding bone marrow tissue  
XX polypeptides of the invention.  
XX  
SQ Sequence 1973 BP; 604 A; 338 C; 352 G; 679 T; 0 other:  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-6 x AAS96223 ..  
  
Align seg 1/1 to: AAS96223 from: 1 to: 1973  
  
27 SerArgAspLeuAlaIysArgIu 34  
|||||  
1683 TCTAGGATCTACGCAAGCGAGG 1706  
  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAA71645  
  
seq\_documentation\_block:  
ID AAA71645 standard; DNA; 2052 BP.  
XX  
AC AAA71645;  
XX  
DF 14-DEC-2000 (first entry)  
XX  
DE Human aspartate protease psl 5 protein encoding DNA.  
XX  
KW Aspartate protease; neuroprotectant; neurotrophic; gamma-secretase activity;  
KW amyloid precursor protein; APP; beta-amyloid; cleavage; Notch; Irel;  
KW Alzheimer's disease; psl 5; human; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200043505-A2.  
XX  
PD 27-JUL-2000.  
XX

```

PE 19-JAN-2000; 2000MO-EP00390.
XX
XX 22-JAN-1999; 99DE-1002550.
PR 08-JUN-1999; 99DE-1025946.
PR 24-JUN-1999; 99DE-1029115.
XX
PA (MEMO-) MEMOREC MEDICAL MOLECULAR RES COLOGNE ST.
RA (HOEM/) HOFMANN K.
XX
XX Hofmann K;
PI
XX WPI: 2000-514726/46.
DR P-PSDB; AAB10556.
XX
XX New protease that cleaves amyloid precursor protein, useful e.g. for
PT treating Alzheimer's disease, and related nucleic acid, inhibitors and
PT antibodies -
XX
XX Claim 4; Page 31-32; 32pp; German.
XX
XX This invention describes a novel protease (I) that has two aspartate
CC residues in a catalytically active structure. The aspartate residues
CC are present in motifs of formulae X.10X.26D and X.3X.49X.5 both present
CC in a transmembrane region. The products of the invention have
CC neuroprotective and neurotropic activity. (I) represent the previously
CC unidentified gamma-secretase activity that is involved in conversion of
CC amyloid precursor protein (APP) to beta-amyloid, and also in cleavage of
CC other membrane proteins such as Notch (implicated in development of the
CC nervous system); N-terminal signal peptides; C-terminal propeptides;
CC generation of peptides for presentation in the context of major
CC histocompatibility complex molecules, and cleavage of the endoplasmic
CC reticulum stress sensor protein Irel. (I), also nucleic acid (II) that
CC encodes it, and its inhibitors and antibodies (Ab) are used to treat or
CC diagnose diseases that are associated with cleavage of amyloid precursor
CC protein, especially Alzheimer's disease, aberrant breakdown of
CC hydrophobic signal peptides, or accumulation of unfolded proteins in the
CC endoplasmic reticulum. They can also be used to modify presentation of
CC hydrophobic peptides by histocompatibility complex molecules, in cases of
CC e.g. viral infection, cancer and transplant rejection. This sequence
CC encodes the human aspartate protease psl 5 which is described in the
CC method of the invention.
XX
XX Sequence 2052 BP; 359 A; 699 C; 576 G; 418 T; 0 other;
S0
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-973-363-6 x AAA71645/rev ..
Align seg 1/1 to reverse of: AAA71645 from: 1 to: 2052
33 ArgGluAlaGlnArgLeuCysGly 40
|||||
959 AGAGAGGCCCGACAGAGCTGTGCGGA 936
seq_name: /SIDSI/gcggdata/hold-geneseq/geneseqn-emb1/AA2001A.DAT:AAK84450
seq_documentation_block:
ID AAK84450 standard; DNA; 5271 BP.
AC AAK84450;
XX
XX 07-NOV-2001 (first entry)
DT
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39262.
XX
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX

```

OS	homo sapiens.	
XX	WO200157182-A2.	
PN		
XX		
XX	09-AUG-2001.	
PD		
XX	17-JAN-2001.	2001WO-US01354
PE		
XX		
PR	31-JAN-2000.	2000US-0179065
PR	04-FEB-2000.	2000US-0180628
PR	24-FEB-2000.	2000US-0184664
PR	02-MAR-2000.	2000US-0186350
PR	16-MAR-2000.	2000US-0189874
PR	17-APR-2000.	2000US-0190076
PR	18-APR-2000.	2000US-0198123
PR	19-MAY-2000.	2000US-0205515
PR	07-JUN-2000.	2000US-0209467
PR	28-JUN-2000.	2000US-0214886
PR	30-JUN-2000.	2000US-0215135
PR	07-JUL-2000.	2000US-0216647
PR	07-JUL-2000.	2000US-0216880
PR	11-JUL-2000.	2000US-0217487
PR	11-JUL-2000.	2000US-0217496
PR	14-JUL-2000.	2000US-0218290
PR	26-JUL-2000.	2000US-0220963
PR	26-JUL-2000.	2000US-0220964
PR	14-AUG-2000.	2000US-0224518
PR	14-AUG-2000.	2000US-0224519
PR	14-AUG-2000.	2000US-0225213
PR	14-AUG-2000.	2000US-0225266
PR	14-AUG-2000.	2000US-0225267
PR	14-AUG-2000.	2000US-0225368
PR	14-AUG-2000.	2000US-0225370
PR	14-AUG-2000.	2000US-0225477
PR	14-AUG-2000.	2000US-0225547
PR	14-AUG-2000.	2000US-0225758
PR	14-AUG-2000.	2000US-0225759
PR	18-AUG-2000.	2000US-0226279
PR	22-AUG-2000.	2000US-0226681
PR	22-AUG-2000.	2000US-0226688
PR	22-AUG-2000.	2000US-0227182
PR	30-AUG-2000.	2000US-0227204
PR	01-SEP-2000.	2000US-0229287
PR	01-SEP-2000.	2000US-0229343
PR	01-SEP-2000.	2000US-0229345
PR	01-SEP-2000.	2000US-0229345
PR	05-SEP-2000.	2000US-0229309
PR	05-SEP-2000.	2000US-0229513
PR	06-SEP-2000.	2000US-0230437
PR	06-SEP-2000.	2000US-0230438
PR	08-SEP-2000.	2000US-0231242
PR	08-SEP-2000.	2000US-0231243
PR	08-SEP-2000.	2000US-0231424
PR	08-SEP-2000.	2000US-0231444
PR	08-SEP-2000.	2000US-0231444
PR	08-SEP-2000.	2000US-0231444
PR	08-SEP-2000.	2000US-0231480
PR	08-SEP-2000.	2000US-0232081
PR	12-SEP-2000.	2000US-0231968
PR	14-SEP-2000.	2000US-0232397
PR	14-SEP-2000.	2000US-0232398
PR	14-SEP-2000.	2000US-0232399
PR	14-SEP-2000.	2000US-0232400
PR	14-SEP-2000.	2000US-0232401
PR	14-SEP-2000.	2000US-0233063
PR	14-SEP-2000.	2000US-0233064
PR	14-SEP-2000.	2000US-0233064
PR	21-SEP-2000.	2000US-0234223
PR	21-SEP-2000.	2000US-0234274
PR	25-SEP-2000.	2000US-0234977
PR	25-SEP-2000.	2000US-0234998
PR	26-SEP-2000.	2000US-0235484

27-SEP-2000: 2000US-0235834.  
 PR 27-SEP-2000: 2000US-0235836.  
 PR 29-SEP-2000: 2000US-0236327.  
 PR 29-SEP-2000: 2000US-0236367.  
 PR 29-SEP-2000: 2000US-0236368.  
 PR 29-SEP-2000: 2000US-0236369.  
 PR 29-SEP-2000: 2000US-0236370.  
 PR 02-OCT-2000: 2000US-0236802.  
 PR 02-OCT-2000: 2000US-0237037.  
 PR 02-OCT-2000: 2000US-0237038.  
 PR 02-OCT-2000: 2000US-0237039.  
 PR 02-OCT-2000: 2000US-0237040.  
 PR 13-OCT-2000: 2000US-0239945.  
 PR 13-OCT-2000: 2000US-0239947.  
 PR 20-OCT-2000: 2000US-0240960.  
 PR 20-OCT-2000: 2000US-0241221.  
 PR 20-OCT-2000: 2000US-0241785.  
 PR 20-OCT-2000: 2000US-0241786.  
 PR 20-OCT-2000: 2000US-0241787.  
 PR 20-OCT-2000: 2000US-0241808.  
 PR 20-OCT-2000: 2000US-0241809.  
 PR 01-NOV-2000: 2000US-0244617.  
 PR 08-NOV-2000: 2000US-0246474.  
 PR 08-NOV-2000: 2000US-0246475.  
 PR 08-NOV-2000: 2000US-0246476.  
 PR 08-NOV-2000: 2000US-0246477.  
 PR 08-NOV-2000: 2000US-0246478.  
 PR 08-NOV-2000: 2000US-0246523.  
 PR 08-NOV-2000: 2000US-0246524.  
 PR 08-NOV-2000: 2000US-0246525.  
 PR 08-NOV-2000: 2000US-0246526.  
 PR 08-NOV-2000: 2000US-0246527.  
 PR 08-NOV-2000: 2000US-0246528.  
 PR 08-NOV-2000: 2000US-0246529.  
 PR 08-NOV-2000: 2000US-0246530.  
 PR 08-NOV-2000: 2000US-0246531.  
 PR 08-NOV-2000: 2000US-0246532.  
 PR 08-NOV-2000: 2000US-0246533.  
 PR 17-NOV-2000: 2000US-0249207.  
 PR 17-NOV-2000: 2000US-0249208.  
 PR 17-NOV-2000: 2000US-0249209.  
 PR 17-NOV-2000: 2000US-0249210.  
 PR 17-NOV-2000: 2000US-0249211.  
 PR 17-NOV-2000: 2000US-0249212.  
 PR 17-NOV-2000: 2000US-0249213.  
 PR 17-NOV-2000: 2000US-0249214.  
 PR 17-NOV-2000: 2000US-0249215.  
 PR 17-NOV-2000: 2000US-0249216.  
 PR 17-NOV-2000: 2000US-0249217.  
 PR 17-NOV-2000: 2000US-0249218.  
 PR 17-NOV-2000: 2000US-0249244.  
 PR 17-NOV-2000: 2000US-0249245.  
 PR 17-NOV-2000: 2000US-0249264.  
 PR 17-NOV-2000: 2000US-0249265.  
 PR 17-NOV-2000: 2000US-0249297.  
 PR 17-NOV-2000: 2000US-0249299.  
 PR 17-NOV-2000: 2000US-0249300.  
 PR 01-DEC-2000: 2000US-0250160.  
 PR 01-DEC-2000: 2000US-0250391.  
 PR 05-DEC-2000: 2000US-0251030.  
 PR 05-DEC-2000: 2000US-0251988.  
 PR 05-DEC-2000: 2000US-0256719.  
 PR 06-DEC-2000: 2000US-0251479.  
 PR 08-DEC-2000: 2000US-0251868.  
 PR 08-DEC-2000: 2000US-0251869.  
 PR 08-DEC-2000: 2000US-0251989.  
 PR 11-DEC-2000: 2000US-0251990.  
 PR 05-JAN-2001: 2001US-0259678.  
 XX  
 PA (HOMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-483426/52.  
 XX  
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 PS  
 PS Disclosure; SEQ ID NO 39262; 3071pp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921 (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC to AAK87694 represent human immune/hematopoietic-derived cells. AAK64703  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 5271 BP; 1866 A; 897 C; 951 G; 1757 T; 0 other;  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-6 x AAK84450 ..  
 Align seg 1/1 to: AAK84450 from: 1 to: 5271  
 27 SerArgAspLeuAlaLysArgIleu 34  
 |||||||||||||||||||||||||  
 4307 TCTAGGAGCTGATGCAAGCGAGAG 4330

.....

---







17 rargalaasptyrleuileylsleuleuserarapsleuala1ays 32  
|||||  
328 CCGTGCAGACTACCTCATCAATTAATTACTAGTAGAGATCTTCACAAA 283

seq\_name: gb\_est1:AU125712

seq\_documentation\_block:  
LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.  
ACCESSION AU125712  
VERSION AU125712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 866)  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isogai,T.  
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
,Y., Sugano,S., Isogai,T.)  
Unpublished (2000)  
JOURNAL Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix  
Research Institute: cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
Location/Qualifiers  
FEATURES  
source 1..866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NT2RM4002061"  
/clone\_lib="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal  
precursor cells"  
BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 32.00 Length: 32  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-6 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

seq\_documentation\_block:  
LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',

ACCESSION mRNA sequence.  
VERSION BE895133  
KEYWORDS BE895133.1 GI:10358221  
SOURCE EST.  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1028)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: ATCC/DCPD/DPF  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINTL at:  
<http://image.jnl.gov>  
Plate: LLM9753 Row: h Column: 16  
High quality sequence stop: 488.  
Location/Qualifiers  
FEATURES  
source 1..1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:3921087"  
/clone\_lib="NIH-MGC-72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-Sport6; Site-1: NotI;  
Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life  
Technologies."  
BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

alignment\_scores:  
Quality: 32.00 Length: 32  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-6 x BE895133 ..

Align seg 1/1 to: BE895133 from: 1 to: 1028

seq\_documentation\_block:  
LOCUS BE239967 821 bp mRNA linear EST 14-NOV-2000  
DEFINITION 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
mRNA sequence.  
ACCESSION BE239967  
VERSION BE239967.1 GI:11153890  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 821)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE



## JOURNAL

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9abbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

## FEATURES

High quality sequence stop: 562.

## source

1 821

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_image="413129"

/clone\_id="NHLKCC-34"

/tissue\_type="from chronic myelogenous leukemia"

/lab\_host="DH10B (TI phage-resistant)"

/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggccgcttggcc); Site\_2: SfiI (ggccattatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGCGCATTTATGCGC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGAGCGCGGACG-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

## BASE COUNT

345 a 149 c 190 g 137 t

## ORIGIN

## alignment\_scores:

Quality: 27.00 Length: 27  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-6 x BF239967 ..

Align seg 1/1 to: BF239967 from: 1 to: 821

6 Proasplyslysproglnalalysglnleuglntharqalaastrytie 22  
|||||  
3 CCCGATAAAAACCAACAGCAACAGCTGCAGACCCGCGACACTACT 52  
|||||  
22 uilelslsleuuseraryaspleuualys 32  
|||||  
53 CATCAATTACTAGTAGAGACTTGCAAAA 83

seq\_name: gb\_est1:AL659353

## seq\_documentation\_block:

LOCUS AL659353 593 bp mRNA linear EST 13-DEC-2001  
DEFINITION AL659353 XGC-neurula silurana tropicalis cDNA clone TNeu045e20 5', mRNA sequence.

## ACCESSION

AL659353  
AL659353.1 GI:17672995

## KEYWORDS

EST  
western clawed frog  
silurana tropicalis  
Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:  
Amphibia: Batrachia: Anura: Mesobatrachia: Pipidae: Pipidae;  
Xenopodinae: Silurana.

## REFERENCE

1. (bases 1 to 593)  
Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.  
Sanger Xenopus tropicalis EST project 2001 (10-2001)

## JOURNAL

Unpublished (2001)  
Contact: Huckle E  
Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: tropes@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TNeu045e20.sp6

Sequencing primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.

Location/Qualifiers

1 593

/organism="Silurana tropicalis"

/db\_xref="taxon:8364"

/clone="TNeu045e20"

/clone\_id="XGC-neurula"

/dev\_stage="neurula"

/lab\_host="Escherichia coli DH10B"

/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dT primed from 5' end of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

## BASE COUNT

119 a 134 c 114 g 226 t

## ORIGIN

## alignment\_scores:

Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-6 x AL659353/rev ..

Align seg 1/1 to reverse of: AL659353 from: 1 to: 593

1 IileuProaspaspProasplyslysproglnalalysglnleuglnth 17  
|||||  
346 ATTTTACGATGATCCAGATAGAAAGCCCAACGACGACGACTACGAC 257  
|||||  
17 TATGAAATATATATATATATATATATATATATATATATATATAT 26  
|||||  
296 CAGAGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 269  
|||||

seq\_name: gb\_est1:AL644594

## seq\_documentation\_block:

LOCUS AL644594 645 bp mRNA linear EST 07-NOV-2001  
DEFINITION AL644594 XGC-egg silurana tropicalis cDNA clone LIEd12 5', mRNA sequence.

## ACCESSION

AL644594  
AL644594.1 GI:16796719

## KEYWORDS

EST  
western clawed frog  
silurana tropicalis  
Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:  
Amphibia: Batrachia: Anura: Mesobatrachia: Pipidae: Pipidae;  
Xenopodinae: Silurana.

## REFERENCE

1. (bases 1 to 645)  
Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.  
Sanger Xenopus tropicalis EST project 2001 (10-2001)

## JOURNAL

Unpublished (2001)  
Contact: Huckle E  
Sanger Centre

## FEATURES

source

1. 645

/organism="Silurana tropicalis"

/db\_xref="taxon:8364"

/clone="LIEd12"

```
/clone_1lb="Xgc-egg"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/notes="Vector: pCS107. Site_1: EcoRI; Site_2: NotI; CDNA
was oligo dT primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut CDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"

BASE COUNT      222 a      125 c      156 g      141 t      1 others
ORIGIN

alignment_scores:
  Quality:      26.00      Length:      26
  Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-08-973-363-6 x AL644594 ..

Align seg 1/1 to: AL644594 from: 1 to: 645

1 11leuproaspaspProaspLysProGlnAlaLysGlnLeuGlnTh 17
|||||
456 ATTTTACCAATGATCCAGATAGAGACCCCAAGCCAGCAGCTACAGAC 505
|||||
17 rArgAlaAspTyrIleuIleLysLeuLeu 26
|||||
506 CAGAGCTGACTACCTCATTAACCTTCTC 533
|||||

seq_name: gb_est1:AM97058

seq_documentation_block:
LOCUS      AM97058      686 bp      mRNA      linear      EST 05-JUN-2000
DEFINITION OV3-BN0047-150400-152-c03 BN0047 Homo sapiens CDNA, mRNA sequence.
ACCESSION  AM97058
VERSION     AM97058.1 GI:8257292
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 686)
            Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV3-BN0047-150
            400-152-c03&ts=2000-04-15&tl=1)
            Seq primer: puc 18 forward
            High quality sequence start: 19
            High quality sequence stop: 678.
            Location/Qualifiers
                1..686
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_1lb="BN0047"
                /dev_stage="Adult"
```

```
/note="Organ: breast_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESPEs PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and CDNA amplification were performed under
low stringency conditions."

BASE COUNT      165 a      154 c      126 g      241 t
ORIGIN

alignment_scores:
  Quality:      25.00      Length:      25
  Ratio:        1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-08-973-363-6 x AM97058/rev ..

Align seg 1/1 to reverse of: AM97058 from: 1 to: 686

9 LysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrIleLysIle 25
|||||
104 AACCCAGACAGCAAAAGTTCAGACCCGTGCAGACATCTCATCAAAATT 55
|||||
25 ulUserArgAspLeuAlaLysArg 33
|||||
54 ACTTAGTAGAGATCTTCGCAAAAAGA 30
|||||

seq_name: gb_est1:AM96787

seq_documentation_block:
LOCUS      AM96787      337 bp      mRNA      linear      EST 05-JUN-2000
DEFINITION OV3-BN0047-230200-102-d03 BN0047 Homo sapiens CDNA, mRNA sequence.
ACCESSION  AM96787
VERSION     AM96787.1 GI:8257021
KEYWORDS   EST.
SOURCE      human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 337)
            Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE    20202663
COMMENT    Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-QV3-BN0047-230
            200-102-d03&ts=2000-02-23&tl=1)
            Seq primer: puc 18 forward
            High quality sequence start: 2
            High quality sequence stop: 337.
            Location/Qualifiers
                1..337
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_1lb="BN0047"
                /dev_stage="Adult"
```

```

/Note="Organ: breast-normal; Vector: puc18; Site_1: Sma1;
Site_2: Sma1. A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      83 a      59 c      72 g      123 t
ORIGIN

```

```
alignment_scores:
  Quality: 23.00
  Ratio: 1.000
  Percent Similarity: 100.000
  Length: 23
  Gaps: 0
  Percent Identity: 100.0000
```

alignment\_block:  
US-08-973-363-6 x AW996787/rev

Align seg 1/1 to reverse of: AW996787 from: 1 to: 337

```

1 11eleuProaspaspProaspLysLysProglinalalysglInleuglntH 17
    |||||
105 ATTCTTCAGATGATCCCGATAAAAAACACACAAAGCAAAACAGTTGCAGAC 56

```

17 rargalaaspyrleuile 23  
|||  
55 CCGTGACAGACTATCTCATC 37

seq\_name: gb\_est1:BB830730

seq\_documentation\_block:

LOCUS	438 bp	mRNA	linear	EST 19-NOV-2001
BB830730				
BB830730	RIKEN full-length enriched, mammary gland			RCB-0527
JY9-MC(B)	cDNA Mus musculus cDNA clone G930013K04 3', mRNA			

ACCESSION	BB830730
VERSION	BB830730.1
REVISION	GI:17008973

SOURCE ORGANISM	house mouse Mus musculus
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus 1 (bases 1 to 438)

Akimura, T., Arikawa, T., Carrinck, P., Furuno, M., Hanagaki, T.,

Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Konda, M., Matsuyama, T., Nakamura, M., Nisii, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Nakamura, M., Nisii, K., Nomura, K., Numasaki, R., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakzunov, N., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, H., Takahashi, F., Takaku-Akahori, S., Tanaka, T., Tomaru, A., Toyota, T., Watanabe, A., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al. 2001)

COMMENT Contact: Yoshihide

Genome Research Center (GRC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suhei-ro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel.: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsic.riken.go.jp,  
URI: <http://genome.gsic.riken.go.jp/>  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh,  
, M., Komou, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
Yagci, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsutani,  
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.

RIKEN Integrated sequence analysis (RISA) system-384-Format Resequencing pipeline with 384 multiplexed sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.ysc.riken.go.jp>) for further details.

e mouse tissues.

-FEATURES	Location/Qualifiers
Source	1. .438

```

/clone.lib=RKEN full-length enriched, mammary gland
RCB-0527 yfg-MC(B) CDNA"
/tissue_type="mammary gland"
/cell_line="RCB-0527 yfg-MC(B)"
BASE COUNT      138 a      96 c      108 g      96 t
ORIGIN

```

alignment\_scores:

Quality:	22.00	Length:	22
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

```
alignment_block:
US-08-973-363-6 x BB830730
```

Align seg 1/1 to: BB830730 from: 1 to: 438

1 11leuProaspaspProaspPlystLysPProGlnAlaLysGlnLeuGlnTh 17  
 |||||  
 300 ATTCTTCCAGATGATCCGTGATAAANAACCAAGCAAAACAGTTACAGAC 34

```

17  rargAlaAspTyrLeu 22
    |||||
350 CCGTGCAGACTACCTC 365

```

seq\_name: gb\_est1:BB834922

seq_documentation_block:			
LOCUS	BB834922	446 bp	mRNA linear
DEFINITION	BB834922	RIKEN full-length enriched,	EST 19-NOV-2001
		mammary gland	RCB-0527

sequence.

VERSION	BB834922.1	GI:17013165
KEYWORDS	EST.	
SOURCE	house mouse.	

## References

REFERENCE  
1 (bases 1 to 446)

EXHIBIT 101

**JOURNAL COMMENT**

**TITLE**

**RIKEN Encyclopedia of Mouse Full-Length cDNAs (Akimura, T., et al. 2001)**

Unpublished (2001)

Contact: Yoshinobu Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsukubumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsic.riken.go.jp,  
 URL: http://genome-gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, M., Sugahara, Y., Shibata, K., Itoh  
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura  
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and  
 Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
 Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome-gsc.riken.go.jp>) for  
 further details.  
 e mouse tissues.

FEATURES  
 source  
 1. 446  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="G930033j21"  
 /clone\_lib="RIKEN full-length enriched, mammary gland  
 RCB-0527 Jy9-MC(B) cDNA"  
 /tissue\_type="mammary gland"  
 /cell\_line="RCB-0527 Jy9-MC(B)"

BASE COUNT  
 139 a 99 c 108 g 100 t

ORIGIN

alignment\_scores:  
 Quality: 22.00 Length: 22  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-6 x BB834922 ..

Align seg 1/1 to: BB834922 from: 1 to: 446

1 11leuProAspPProAspLysLysPProGlnAlaLysGlnLeuGlnThr 17  
 |||  
 309 ATTCTTCACAGATGATCCTGATMAAACACACAGCAACAGTTACAGAC 358  
 17 TARGAlAsPTyrlleu 22  
 |||  
 359 CCGTGCAGACTACTCTC 374

seq\_name: gb\_gss:CNS04DVG

seq\_documentation\_block:  
 LOCUS CNS04DVG 856 bp DNA linear GSS 21-MAY-2000  
 DEFINITION Tetradon nigriviridis genome survey sequence 17 end of clone  
 103008 of library G from Tetradon nigriviridis, genomic survey  
 sequence.  
 ACCESSION AL286261 GI:8024707  
 VERSION AL286261.1  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetradon nigriviridis.  
 ORGANISM Tetradon nigriviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetraodon.  
 1 (bases 1 to 856)  
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fitzames, C., Fisher, C.,

Bonneau, L., Billault, A., Quetier, F., Saurin, M., Bernot, A. and  
 Weissenbach, J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetradon nigriviridis  
 Unpublished  
 2 (bases 1 to 856)  
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C.,  
 Bernot, A., Fitzames, C., Winkler, P., Brothier, P., Quetier, F.,  
 Saurin, M. and Weissenbach, J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetradon nigriviridis DNA sequence  
 Unpublished  
 3 (bases 1 to 856)  
 Genoscope.  
 Direct Submission  
 Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the tetradon nigriviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetradon>.

FEATURES  
 source  
 1. 856  
 /organism="Tetradon nigriviridis"  
 /db\_xref="taxon:99883"  
 /clone="103K08"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0B6103BF04LP1-end : T7"

BASE COUNT  
 84 a 308 c 252 g 206 t 6 others

ORIGIN

alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-6 x CNS04DVG/rev ..

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

2 leuProAspPProAspLysLysPProGlnAlaLysGlnLeuGlnThr 18  
 |||  
 674 CTGCCGATGATCCTGATMAAGCCTCAGGCCCAACAGTTACAGACCAG 625  
 18 GAlA 19  
 |||  
 624 AGCC 621

seq\_name: gb\_est2:BF881342

seq\_documentation\_block:  
 LOCUS BF881342 129 bp mRNA linear EST 17-JAN-2001  
 DEFINITION PM0-ET0208-031200-001-f12 ET0208 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BF881342  
 VERSION BF881342.1 GI:12271468  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 129)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M. R.,  
 Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
 Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,  
 Brustein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare  
 M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
 Simpson, A. J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL:  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM06t2-PM0-ET0208-  
031200-001-f12xt3-2000-12-03&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 22  
High quality sequence stop: 128.  
Location/Qualifiers

# FEATURES

## source

1. 129  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="ET0208"  
/dev\_stage="Adult"  
/note="Organ: lung-tumor; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 21 a 48 c 23 g 37 t  
ORIGIN

## alignment\_scores:

Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-6 x BF881342/rev ..

Align seg 1/1 to reverse of: BF881342 from: 1 to: 129

13 lysglnleuglnthrargAlAspTyleu 22  
|||||  
108 AAGCAGCTACAGACCGAGCGGATTCCTTG 79

**THIS PAGE BLANK (USPTO)**



COMMENT Other publication AU 5906996 961224.  
FEATURES Location/Qualifiers  
source 1..6608 /organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN  
alignment\_scores:  
Quality: 41.00 Length: 41  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-7 x A58691 ..  
Align seg 1/1 to: A58691 from: 1 to: 6608  
1 lleuProaspProaspPlysLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAAGCCAGCAAGCCTACAGAC 4129  
17 rArgAlaAspTyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
4130 CCGTGCAGACTACTCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 4179  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
4180 AAGCAAAAGGCTTGCTGTGCA 4202  
seq\_name: gb\_ov:AF004397  
seq\_documentation\_block:  
LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
DEFINITION Gallus gallus chemo-helicase-DNA-binding on the Z chromosome  
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete  
cvs.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 6872)  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
REFERENCE 2 (bases 1 to 6872)  
AUTHORS Griffiths,R. and Korn,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK  
FEATURES  
source Location/Qualifiers  
1..6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1..6872  
/gene="CHD-Z"  
228..5654  
/gene="CHD-Z"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chromo-helicase-DNA-binding on the Z chromosome  
protein"  
/protein\_id="AAC60282.1"  
/translation="MNGHSDSESVNRSSSESRDDSSAGSGSGSSGSSGSSGSS  
SOSGSSDSESGSESSESSTREKKOYAKPKADGSEFWKSSPSIIAVQRSVA

LKQOQOQAAASDSSGSESSSEDSADSSSETKKKKKHNDQMOMSGSGSVSGTGS  
DSEADGDKSSCESESDYEPKKNKYSKPSRIKPKSGKSTGQKKQLDSESEE  
DDDEDYDKRSROATVNVSYKEAETKTDSDLLVECEDEVPTQEDDEFETIKEMD  
SRTGRGATGASTTITVAVEDGDDPNAGPEKSKSEIEIOYLTKMKGSHHNWTEET  
LKQNVKGNKRLDNTKKKQDETKRNLKANSPEDEVYINCOQLTDLHKQYIVERTI  
AHSNQKSAAGIPDYTCRWGLPYSECSWEDGALIAKFFQARIDETFSRNQSTTPPKD  
CIVLQRPREFVALKROPYSIGHESLELDYOLNLMVLASHMCKGNSCTIADENGSLG  
KTIOISFLNLYLFEHQLYGPFLRLRYPLSTLSWQREIQTMAPQNAVYLDITSRN  
MIRTHEMHMPOTKRLKENILLTYETELIKDKSFGLGNAPFGVDEAHLKNDLSLY  
RTLIDKSNHRLITGTPLONSLKELMSLHIFIMPEKSSMEDFEDEHGKGRGYAS  
LKELEPELILRRYKDDVEKSLPAKYEQILRMMSALQOYIYWIITIRYALKSKSGS  
STSGPLNTIMELKCCNHCYLIRPPDNEFYKQELQHLISSGKLITLDLILRL  
ERGNRLIFSQWRLMIDIAEYLKROFPFOLDGSIKELKROALDHFNAGSBDPC  
FLLSFRAGLGGINLASDVTYIFSDSDMNDLOAQRARHARGKQKQVYIYLVYKGS  
VEEDILERRAKKKKMDHLYIORMDTGKTVLTGTSPPSSPTPNEESALILPGAE  
LEKPEGEPOEPMIDILIKRAETRENEPGLYVQELLISOFVANSNDDEDIE  
LEPERSRMWEITIPESORRIIEBERQKELEIYMLPRMCAKOISNGBEGRSR  
SRRYSGSDSDSITERRRPRKRPRTIPRENTKGSFDEIRFIRIKSYKFGSLERLD  
AVARDAELVDKSETDLRLGELVHNGCALKLDNSSGQERAGRGLGKVGPPYRISGV  
QVNAKLIVISHEELAPLHKISIPDEERKRYIIPCHTKAHHIDIMGKEDSDNLVGI  
YVYGGSWEMIKMDPLSLTOKILPDDPKPKQAKOLOTRADYLIKLKLDAREAO  
RLAGANSKRRTKTRKKNKMAKSKIKETIKSSPOPSKSPDEDEEDENKDEIYSVK  
HIAKIKITKEENBEKPEPDIGIKKAEKERREKREKREKREKREKREKREKREK  
NKEKENKVESTOKKEVEKERNVMSSENKSKKIPILDTPVHITATSPVISE  
ESEELHQKTFSVCKEMRPVKALQOLRPERGLESERDQLEHTROCLIKIGHITFCL  
KEYTDEQIKQWRKMLMIFVSFTFEDAKRLIKLYKHAIKKROESQOHDONISSNV  
TVYIRNPDIYERLKEETNHDSSRDSYSPRHLISYDHHKRDHODGDAYKSPSRPY  
SAFNSGKHDRMDHYKODSRYSDSKHKRLDHRDRHDSNLEGLNLRGSHDSHRSH  
SDPHRHSDRHSSEYSHKSSRDYRYHSDMWDHNASGSGSPSPIDQSRPYSRPLG  
HRSPEHSSDHRKSTPEHTWSSKKT"  
4341..4604  
/gene="CHD-Z"  
/note="short insert found in longer variant mRNA of CHD-Z"  
BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN  
alignment\_scores:  
Quality: 41.00 Length: 41  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-7 x AF004397 ..  
Align seg 1/1 to: AF004397 from: 1 to: 6872  
1 lleuProaspProaspPlysLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAAAGCCAGCAAGCCTACAGAC 4129  
17 rArgAlaAspTyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
4130 CCGTGCAGACTACTCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 4179  
34 luAlaGlnArgLeuAlaGlyAla 41  
|||||  
4180 AAGCAAAAGGCTTGCTGTGCA 4202  
seq\_name: gb\_pat:A58684  
seq\_documentation\_block:  
LOCUS A58684 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 3 from Patent WO96399505.  
ACCESSION A58684  
VERSION A58684.1 GI:3714247  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN CHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN



BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
ISIS INNOVATION (GB)  
COMMENT Other publication AU 5906996 961224.  
FEATURES Location/Qualifiers  
source 1. 153  
/db\_xref="taxon:32644"  
BASE COUNT 58 a 40 c 31 g 24 t  
ORIGIN

alignment\_scores:  
Quality: 34.00 Length: 34  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x A58684 ..  
Align seg 1/1 to: A58684 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAGAAACCCGACGCAAGCAGCTACAGCCGTCGACGACTCATTA 101  
|||||  
24 sleuLeuAsnLysAspLeuAlaArgLysGluAlaGlnArgLeuAlaGly 41  
|||||  
102 ATTACTGAATTAAGACCTTGCAAGAAAGACCAAGCAAGCTTCTGCTG 151  
41 la 41  
152 CA 153

seq\_name: gb\_pat:A58685

seq\_documentation\_block:  
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 4 from Patent WO9639505.  
ACCESSION A58685  
VERSION A58685.1 GI:3714248  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths, R. and Tiwari, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source 1. 153  
/db\_xref="taxon:32644"  
BASE COUNT 56 a 36 c 31 g 30 t  
ORIGIN

alignment\_scores:  
Quality: 34.00 Length: 34  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x A58685 ..  
Align seg 1/1 to: A58685 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAGAAACCCGACGCTACAGCTTACAGCCGTCGACGACTCATTA 101  
|||||

24 sleuLeuAsnLysAspLeuAlaArgLysGluAlaGlnArgLeuAlaGly 41  
|||||  
102 ATTACTGAATTAAGACCTTGCAAGAAAGACCAAGCAAGCTTCTGCTG 151  
41 la 41  
152 CA 153

seq\_name: gb\_pat:A58686

seq\_documentation\_block:  
LOCUS A58686 153 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 5 from Patent WO9639505.  
ACCESSION A58686  
VERSION A58686.1 GI:3714249  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths, R. and Tiwari, B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
Location/Qualifiers  
source 1. 153  
/db\_xref="taxon:32644"  
BASE COUNT 60 a 34 c 31 g 28 t  
ORIGIN

alignment\_scores:  
Quality: 27.00 Length: 27  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x A58686 ..  
Align seg 1/1 to: A58686 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAGAAACCCGACGCAAGCAGCTGACGACCGTCGACGACTCATTA 101  
|||||  
24 sleuLeuAsnLysAspLeuAlaArgLysGlu 34  
|||||  
102 ATTACTGAATTAAGACCTTGCAAGAAAGAA 132

seq\_name: gb\_to:MUSCHD1X

seq\_documentation\_block:  
LOCUS MUSCHD1X 5349 bp mRNA linear ROD 19-SEP-1996  
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.  
ACCESSION L10410 X66028  
VERSION L10410.1 GI:455014  
KEYWORDS DNA binding protein.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 5349)  
AUTHORS Delmas, V., Stokes, D.G. and Perry, R.P.  
TITLE A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)  
MEDLINE 93211972  
REFERENCE 2 (bases 1 to 5349)  
AUTHORS Perry, R.P.

TITLE Direct Submission  
JOURNAL Submitted (08-Apr-1993) Robert Perry, The Fox Chase Cancer Center,  
7701 Burholme Avenue, Philadelphia, PA 19111, USA  
COMMENT On Feb 16, 1994 this sequence version replaced gi:293322.  
FEATURES  
    source  
        1..5349  
            /organism="Mus musculus"  
            /db\_xref="taxon:10090"  
            /cell\_type="plasmacytoma"  
            1..5349  
            /gene="CHD-1"  
            171..5306  
            /gene="CHD-1"  
            /codon\_start=1  
            /product="DNA-binding protein"  
            /protein\_id="AAB08486.1"  
            /db\_xref="GI:455015"  
            /translation="MNGHSDSESVNNGSGESSQSDDCGSASGSGSGSSGSSGSSGSS  
            SOSGSSDSDSGSGSSSESDTSREKRVQAKPKPVGAEFWKSSPSILAVQRAKL  
            RKQPOAQOQPASSNSGSEEDSSSSGSKARRKHNDEDMQSSGSPQLG  
            SDSESEERKSSCDGTESDYEPRKRVSRKPNKSKNGKILGOKRQIDSEDE  
            DDEDYDNKRSSRQATVNVSYKDEEMKTDSDLLVEYGEDVPQPEDEPFTIRVM  
            DCRVGRKATGATTTIYAEVADGDNPAGPERKKEPDIOYLIIKKGMSHIHNTWTEE  
            TLKOONVRGAKKLDNYKKKQDPTKRLKNASPEDEYVYNCQDELIDLHKQIYERI  
            IAHSQKSAAGLPDYCKWQGLPISGCSWEDALISKFYQICIDEYFSHNSQITPK  
            DCKVLKQRPREFVALKRPQSYIGEGHEGLRDYQLNGLWMLASHWCKGNSCLADEML  
            GKTIQTIISFLNLFHEHQLYGFLLVPLSTLTSQREIQTWASQMANAVYIGADINSR  
            NMIRTHEMMHPQTKRLKFNILITTYEILIKDKAFGLMNAIGVDEARLKNDSLL  
            KTLIDFKSNHRLITGTPLONSLKEIMSLHFIEMKPSMEDFESEHGKREYVA  
            SLHKELEPFLIRVKKDYKSLPAVEOLIRREMSALOKOYKATILTRYKALSQSK  
            GSTSGFLINMELKCCNHCYILKPPDNNEFNKQALQHLIRSGKILIDKLILRL  
            REGRNVLIFSQVNRMLDLIAEYLKROFPQRLDGSIKGELRQALDHPNAGSEDF  
            CFLLSTRAGGLGINLASADTVVIFDSQNPQDLQOARAHRIQOKOVNIYRLVYG  
            SVEEDILERRAKKMYLDLVIQRMDTGKTVLHTGAPSSSTPFNKEELISALIKGAE  
            ELFKPEGEDEPOEMDIDELIKRAETHNEBGPISVGDDELSQFKNFNSMDDDI  
            ELEPRNSKNMEIIPEDORRLREEBERKLEIETMLPRMNCQKQISFNSBERRS  
            RSRRYSGSDSISERRKPKRGRPTIPRENIKGSDAEIRFTKSTKRGEPLERL  
            DAIARDAELVKSSETDLRLGELVHNGCVKALKDSSGTEAGGLGVKGTPTPLIG  
            VOVNAKLIYAEDELIPLHKSIPSDPERKQYTIPOHTKAHFIIDWCKEDSNLLIG  
            IYEVYGSWEMIKMPDLSLTHKILPDDPKKPOAKLOTADYLIKLSRLAKREA  
            OLCGAGSGEMIKRTRAKSKAKMSIKVKEIKRSDSPLSEKSDDDDKINDSKRESK  
            DRSKSVSDAPVHITASGEPPYIAESEBELDQKTFSTCKEMRPYKAAKLOLDPK  
            GLSEBELEHTROCLIKIGDHTTECLAKESNEPOLKQWRKMLWTSKFTEDARKLH  
            KLYKAIKRRQESQONSQNSNVATTHYIRNDEMRLKENTNHDSSRDSYSSDLKH  
            QYHDKRKHQGDYSKSDSRKRPYSFSGNGDHRMDHYRQDSRYYSQREKRRLLD  
            HRSRHRSLGGLKDRCHSDHRSDDHRSMDHRSSEHTHKSSRDYRYLSDWQLD  
            HRASSGPRSPIDORSYGSRPFESHSRSTPHTWSSRKT"  
BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN  
alignment\_scores:  
    Quality: 26.00 Length: 26  
    Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-7 x MUSCHDIX ..  
Align seg 1/1 to: MUSCHDIX from: 1 to: 5349  
1 TlaleuTroaSpaPpLoaPlySLySProGlnAlaLySglnLeuGlnth 17  
|||||  
4026 ATCTCTCCAGATGATCTGATATAAAAAACACAAAGCAAGTATACAGAC 4075  
17 TARGAlaAsPTytleuTleuTleu 26  
|||||  
4076 CCGTGCAGACTACTCTCAACTACTT 4103  
seq\_name: gb\_pr:Af006513  
seq\_documentation\_block:  
LOCUS Af006513 5947 bp mRNA linear PRI 27-NOV-1997

DEFINITION Homo sapiens CHD1 mRNA, complete cds.  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428  
KEYWORDS  
SOURCE  
ORGANISM  
    human.  
    Homo sapiens  
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
    Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
    AUTHORS  
        1 (bases 1 to 5947)  
    JOURNAL  
        Woodage,T., Basral,M.A., Baxevanis,A.D., Hieter,P. and Collins,F.S.  
        Characterization of the CHD family of proteins  
        Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)  
MEDLINE  
    97470991  
REFERENCE  
    2 (bases 1 to 5947)  
AUTHORS  
    Woodage,T.  
TITLE  
    Direct Submission  
JOURNAL  
    Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human  
    Genome Research Institute, National Institutes of Health, 49  
    Convent Drive, Bethesda, MD 20892-4442, USA  
FEATURES  
    source  
        1..5947  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /chromosome="5"  
            /map="5q15-21; near WI-5811"  
            1..5947  
            /gene="CHD1"  
            164..5293  
            /gene="CHD1"  
            /codon\_start=1  
            /product="CHD1"  
            /protein\_id="AAB87381.1"  
            /db\_xref="GI:2645429"  
            /translation="MNGHSDSESVNNGSGESSQSDDCGSASGSGSGSSGSSGSSGSS  
            SOSGSSDSDSGSGSSSESDTSREKRVQAKPKPVGAEFWKSSPSILAVQRAKL  
            KKQOQOQOQOQASNSGSEEDSSSDSDSSSEYKRRKHNDEDMQSSGSPQS  
            GSDSESEERKSSCDGTESDYEPRKRVSRKPNKSKNGKILGOKRQIDSEDE  
            DDEDYDNKRSSRQATVNVSYKDEEMKTDSDLLVEYGEDVPQPEDEPFTIR  
            FMDCTIGRKATGATTTIYAEVADGDNPAGPERKKEPDIOYLIIKKGMSHIHNTWEE  
            EETLKOONVRGAKKLDNYKKKQDPTKRLKNASPEDEYVYNCQDELIDLHKQIYV  
            RIASNSQKSAAGPYDYCKWQGLPISGCSWEDALISKFYQICIDEYFSHNSQITPK  
            FKDCKVLKQRPREFVALKRPQSYIGEGHEGLRDYQLNGLWMLASHWCKGNSCLADEM  
            GKTIQTIISFLNLFHEHQLYGFLLVPLSTLTSQREIQTWASQMANAVYIGADINS  
            SRMIRTHEMMHPQTKRLKFNILITTYEILIKDKAFGLMNAIGVDEARLKNDS  
            LLYKTLIDFKSNHRLITGTPLONSLKEIMSLHFIEMKPSMEDFESEHGKREYV  
            VASJKELEPFLIRVKKDYKSLPAVEOLIRREMSALOKOYKATILTRYKALSQSK  
            SGSTSGFLINMELKCCNHCYILKPPDNNEFNKQALQHLIRSGKILIDKLILRL  
            RLREGRNVLIFSQVNRMLDLIAEYLKROFPQRLDGSIKGELRQALDHPNAGSE  
            DFCFLLSTRAGGLGINLASADTVVIFDSQNPQDLQOARAHRIQOKOVNIYRLVY  
            KGSVEEDILERRAKKMYLDLVIQRMDTGKTVLHTGAPSSSTPFNKEELISALIKG  
            AELEFKPEGEDEPOEMDIDELIKRAETHNEBGPISVGDDELSQFKNFNSMDDDI  
            DIELEPRNSKNMEIIPEDORRLREEBERKLEIETMLPRMNCQKQISFNSBERRS  
            RSRRYSGSDSISERRKPKRGRPTIPRENIKGSDAEIRFTKSTKRGEPLERL  
            RLDAIARDAELVKSSETDLRLGELVHNGCVKALKDSSGTEAGGLGVKGTPTPLIG  
            SGVOVNAKLIYAEDELIPLHKSIPSDPERKQYTIPOHTKAHFIIDWCKEDSNLL  
            IGIYEVYGSWEMIKMPDLSLTHKILPDDPKKPOAKLOTADYLIKLSRLAKREA  
            EALSAGSGEMIKRTRAKSKAKMSIKVKEIKRSDSPLSEKSDDDDKINDSKRESK  
            ERSKSVSDAPVHITASGEPPYIAESEBELDQKTFSTCKEMRPYKAAKLOLDPK  
            GLSEBELEHTROCLIKIGDHTTECLAKESNEPOLKQWRKMLWTSKFTEDARKLH  
            KLYKAIKRRQESQONSQNSNVATTHYIRNDEMRLKENTNHDSSRDSYSSDLKH  
            YHDKRKHQGDYSKSDSRKRPYSFSGNGDHRMDHYRQDSRYYSQREKRRLLD  
            RSRDRHSLGGLKDRCHSDHRSDDHRSMDHRSSEHTHKSSRDYRYLSDWQMDH  
            RASSSGPRSPIDORSYGSRPFESHSRSTPHTWSSRKT"  
BASE COUNT 2130 a 1004 c 1243 g 1570 t  
ORIGIN  
alignment\_scores:  
    Quality: 26.00 Length: 26  
    Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:



ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Homo sapiens, clone Rpl1-58M12  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 143079)  
1 (bases 1 to 143079)  
2 (bases 1 to 143079)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckler, R., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A., Chappel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deatello, K., Dewar, K., Domino, M., Doyle, M., Fenesco, J., Ferrelia, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Garfield, S., Grant, G., Hagos, B., Harford, A., Horton, L., Howland, J., Johnson, R., Jones, C., Kann, L., Karas, A., Klein, J., Landers, J., Lebeck, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K., McPherson, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, D., Notman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W., J., Zimmer, A., and Zody, M.  
Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:1407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L5154  
Center clone name: 58\_M12  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q30  
Consensus quality: 140814 bases at least Q20  
Insert size: 144000; agarose-1p  
Insert size: 142179; sum-of-coverage  
Quality coverage: 4.6 in Q20 bases; agarose-1p  
Quality coverage: 4.7 in Q20 bases; sum-of-coverage  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 38820: contig of 38820 bp in length  
\* 38821 38920: gap of 100 bp  
\* 38921 40411: contig of 1491 bp in length  
\* 40412 40511: gap of 100 bp  
\* 40512 43279: contig of 2768 bp in length  
\* 43280 43379: gap of 100 bp  
\* 43380 46905: contig of 3526 bp in length  
\* 46906 47005: gap of 100 bp  
\* 47006 51830: contig of 4825 bp in length  
\* 51831 51930: gap of 100 bp  
\* 51931 62619: contig of 10689 bp in length  
\* 62620 62719: gap of 100 bp  
\* 62720 75408: contig of 12689 bp in length

FEATURES  
source  
1..143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Rpl1-58M12"  
/clone\_lib="RPCR-11 Human Male BAC"  
1..38820  
/clone="assembly-fragment"  
vector\_side:left  
38921..40411  
/clone="assembly-fragment"  
vector\_side:right  
40512..43279  
/clone="assembly-fragment"  
43380..46905  
/clone="assembly-fragment"  
47006..51830  
/clone="assembly-fragment"  
51931..62619  
/clone="assembly-fragment"  
62720..75408  
/clone="assembly-fragment"  
75509..92516  
/clone="assembly-fragment"  
92617..106409  
/clone="assembly-fragment"  
106510..143079  
/clone="assembly-fragment"  
vector\_side:right  
vector\_end:right  
BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN  
-----  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
-----  
alignment\_block:  
US-08-973-363-7 x AC021449 ..  
Align seg 1/1 to: AC021449 from: 1 to: 143079  
1 11leupProaspaspproasp1ylysp1rogl1alalyg1nleug1n17  
116147 ATCTTCGACGATGATCCGATTAACCAACGACAAACAGTTGCAGAC 116196  
17 rargalasp1ylysp1rogl1alalyg1nleug1n17  
116197 CCGTGCAGACTACCTCATCAATTACTT 116224  
seq\_name: gb\_hlg:AC008531  
-----  
seq\_documentation\_block:  
LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
7 ordered pieces.  
ACCESSION AC008531  
VERSION AC008531.3 GI:12830078  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.

```

TITLE                               Sequencing of Human Chromosome 5
JOURNAL                             Unpublished
REFERENCE                           2 (bases 1 to 145659)
AUTHORS                             DOE Joint Genome Institute.
TITLE                               Direct Submission
JOURNAL                             Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
                                     Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT                              On Feb 14, 2001 this sequence version replaced gi:7528342.

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

-----
Project Information
Center Project Name: 369535
Center Clone name: CIR-HSPC_480B11
-----

Summary Statistics
Consensus quality: 139128 bases at least Q40
Consensus quality: 142556 bases at least Q30
Consensus quality: 143744 bases at least Q20
Estimated insert size: 148000; pulse field gel estimation
Estimated insert size: 145050; sum-of-contigs estimation
Quality coverage: 6.4 in Q20 bases; pulse field gel estimation
Quality coverage: 6.2 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
*   1      56174: contig of 56174 bp in length
*     *    56175: gap of unknown length
*     *    56275: gap of unknown length
*     *    100875: 100974: gap of 44600 bp in length
*     *    100975: 100974: gap of unknown length
*     *    113128: 113227: gap of 12153 bp in length
*     *    113328: 118190: contig of 4953 bp in length
*     *    118191: 118290: gap of unknown length
*     *    118291: 118294: gap of 1104 bp in length
*     *    118295: 119794: gap of unknown length
*     *    119795: 122397: contig of 3503 bp in length
*     *    119798: 122397: gap of unknown length
*     *    123398: 123397: gap of unknown length
*     *    123399: 145653: contig of 22262 bp in length.
*
FEATURES
     source
         1..145659
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="5"
             /clone="CIRC-480B11"
             /clone_lib="caltech human BAC library C"
BASE COUNT      42561 a 26309 c 27580 g 48609 t      600 others
ORIGIN
alignment_scores:
    Quality:          26.00      Length:          26
    Ratio:            1.000      Gaps:              0
Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x AC008531 ..

Align seg 1/1 to: AC008531 from: 1 to: 145659

1 TTTCTTCCAGATGATCCCGATTAAACAACCAAGCAAACAGCTTGACAGAC 46819
ATTCCTTCCAGATGATCCCGATTAAACAACCAAGCAAACAGCTTGACAGAC 46819
17 TATGATAAAPTyrLeuIleLysLeuLeu 26

```

```

46820 CCGTCACAGTACCTCATCAATTAAGT 46847
seq_name: gb_hlg:AC091946
seq_documentation_block:
LOCUS      AC091946              193446 bp    DNA             linear   HTG 09-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-36012, *** SEQUENCING IN
PROGRESS   ***, 33 unordered pieces.
ACCESSION  AC091946
VERSION    AC091946.1 GI:14333882
KEYWORDS   HTG, HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 193446)
            DOE Joint Genome Institute.
            Sequencing of Human Chromosome 5
            Unpublished
            2 (bases 1 to 193446)
            DOE Joint Genome Institute.
            Direct Submission
            Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 544799
            Center clone name: RPC1-11_36012
            -----
            Summary Statistics
            Consensus quality: 15767 bases at least Q40
            Consensus quality: 180259 bases at least Q30
            Consensus quality: 184175 bases at least Q20
            Estimated insert size: 204590; agarose-gel estimation
            Estimated insert size: 190246; sum-of-contigs estimation
            Quality coverage: 5.38 in Q20 bases; agarose-gel estimation
            Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1
            1358: contig of 1358 bp in length
            *
            1359
            1458: gap of unknown length
            *
            1459
            2667: contig of 1209 bp in length
            *
            2668
            2767: gap of unknown length
            *
            2768
            3823: contig of 1056 bp in length
            *
            3824
            3923: gap of unknown length
            *
            3924
            5920: contig of 1997 bp in length
            *
            5921
            6020: gap of unknown length
            *
            6021
            7494: contig of 1474 bp in length
            *
            7495
            7594: gap of unknown length
            *
            7595
            9096: contig of 1502 bp in length
            *
            9097
            9196: gap of unknown length
            *
            9197
            10261: contig of 1064 bp in length
            *
            10261
            10360: gap of unknown length
            *
            10361
            12461: gap of 2100 bp in length
            *
            12461
            12560: gap of unknown length
            *
            12561
            14611: contig of 2051 bp in length
            *
            14612
            14711: gap of unknown length
            *
            14712
            16381: contig of 1670 bp in length
            *
            16382
            16481: gap of unknown length
            *
            16482
            17968: contig of 1487 bp in length
            *
            17969
            18068: gap of unknown length
            *
            18069
            20433: contig of 2365 bp in length

```

```
* 20435 20534: gap of unknown length
* 20535 23515: contig of 2981 bp in length
* 23516 23615: gap of unknown length
* 23616 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5982 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57460: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 74132: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 87611: gap of unknown length
* 87612 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92891 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
FEATURES
source 1..193446
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="RP11-36012"
        /clone_11b="RP11 human BAC library 11"
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN
alignment_scores:
    Quality: 26.00      Length: 26
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-08-973-363-7 x AC091946 ..
Align seg 1/1 to: AC091946 from: 1 to: 193446
1 11leuProaspPpAspPlyLysProGlnAlaLysGlnLeuGlnth 17
|||||
87658 ATTCTTCACATGATGCCGATATAAAACCAACGAAACAGTTGCAGAC 87707
|||||
17 rArgAlaAspTyrIleuIleLysLeuLeu 26
|||||
87708 CCGTGCAGACTACTCATCAATTACTT 87735
seq_name: gb_pr:AC026778
```

```
seq_documentation_block:
LOCUS AC026778 195433 bp DNA linear PRI 01-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.
ACCESSION AC026778
VERSION AC026778.4 GI:14277282
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 195433)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 195433)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 195433)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Jun 1, 2001 this sequence version replaced gi:13677045.
COMMENT Draft Sequence Produced by DOE Joint Genome Institute
www.igf.doe.gov
www.fgci.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.2.
STS Content:
WI-13675 G33101
SHGC-58345 G38487
SHGC-103595 G57841.
FEATURES
source 1..195433
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="CTC-428111"
BASE COUNT 62762 a 37302 c 37040 g 58329 t
ORIGIN
alignment_scores:
    Quality: 26.00      Length: 26
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000
alignment_block:
US-08-973-363-7 x AC026778/rev ..
Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433
1 11leuProaspPpAspPlyLysProGlnAlaLysGlnLeuGlnth 17
|||||
29732 ATTCTTCACATGATGCCGATATAAAACCAACGAAACAGTTGCAGAC 29683
|||||
17 rArgAlaAspTyrIleuIleLysLeuLeu 26
|||||
29682 CCGTGCAGACTACTCATCAATTACTT 29655
seq_name: gb_pr:AC022121
seq_documentation_block:
LOCUS AC022121 219258 bp DNA linear PRI 30-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone CYD-2007H13, complete sequence.
ACCESSION AC022121
VERSION AC022121.6 GI:15375145
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 219258)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission  
Unpublished  
2 (bases 1 to 219258)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 219258)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission  
Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 219258)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
Direct Submission  
Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Aug 30, 2001 this sequence version replaced gl:15148108.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.4.  
STS Content:  
WI-5811 G04974  
WI-13675 G23101  
SHGC-58345 G38487  
SHGC-103595 G57841

FEATURES

source  
1..219258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2007H13"  
BASE COUNT 71954 a 42062 c 40933 g 64309 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AC022121/rev

Align seg 1/1 to reverse of: AC022121 from: 1 to: 219258

1 HleauProaAspProAspLysLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
89450 ATCTTCAGATGATCCGATGAAAACCAACCAACACTTCAGAC 89401  
17 rAgaLaasPTyrLeuLleLysLeuLeu 26  
|||||  
89400 CCGTCAGACATCACTCAATTAATT 89373

**THIS PAGE BLANK (USPTO)**





1188 TCGTGCAGATTACCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 1237  
 34 luAlaGlnArGLeuAlaGlyAla 41  
 |||||||  
 1238 AACGACAGAGACTTGCTGTGCA 1260

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAr42751

seq\_documentation\_block:

ID AAr42751 standard; cDNA; 6608 BP.

AC AAr42751;

DT 12-MAR-1997 (first entry)

DE Chicken CHD-1A gene.

KM Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

KW CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

OS Gallus sp.

FT Key location/Qualifiers

FT CDS 228..5390

FT /\*tag= a

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determin. and to control sex of progeny

XX Claim 1; Fig 5; 76pp; English.

XX The chicken CHD-W gene (AAr42754) acting alone or in conjunction with

CC the closely related CHD-1A gene (AAr42751) is suggested to initiate

CC female development in birds. The sequence of CHD-1A was deduced

CC from 3 clones isolated from a stage 10-12 chicken embryo cDNA

CC library using a great tit CHD-W sequence (see also AAr42755) as probe.

CC The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1

CC chromosome (see also AAr42756-57). It is located on an autosome or X

CC chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-

CC specific signal on hybridisation to genomic DNA of a non-fertile

CC bird and can be used for sex determin. of a bird. CHD-1A nucleic

CC acids can also be used to control the sex of the progeny of a bird.

XX Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores:

Quality: 41.00 Length: 41

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AAr42751 ..

Align seg 1/1 to: AAr42751 from: 1 to: 6608

1 lIleuProAspProAspLysProGlnAlaLysGlnLeuGlnH 17  
 |||||||  
 4080 ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGACGCTACAGAC 4129

17 rArgAlaSPtyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34  
 |||||||  
 4130 CCGTGCAGACTACTCCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 4179  
 34 luAlaGlnArGLeuAlaGlyAla 41  
 |||||||  
 4180 AACGACAAAGCCTTGCTGTGCA 4202

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAr42757

seq\_documentation\_block:

ID AAr42757 standard; DNA; 153 BP.

AC AAr42757;

DT 12-MAR-1997 (first entry)

DE Chick CHD-1A gene fragment.

KM Bird; sex determination; chromodomain-Helicase-DNA binding 1;

KW CHD-1A; CHD-W; W chromosome; ss.

OS Gallus sp.

FT Key location/Qualifiers

FT misc\_difference 52..81

FT /\*tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51

FT and are ignored in the translated amino

FT acid sequence given in Fig 3"

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

P-PSDB; AAm08147.

XX Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determin. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAr42756) of the mouse CHD-1 gene show homology

CC to portions of the chicken CHD-1A (A = Avian) gene (AAr42757),

CC chicken CHD-W (W refers to the W chromosome) gene (see also AAr42758)

CC and the great tit CHD-W gene (see also AAr42759). Translated amino

CC acid sequences of this region are provided in AAm08146-49. The

CC CHD-1A (see also AAr42751) and CHD-W (see also AAr42754-55) genes

CC determine sex in birds and can be used to identify the sex of an

CC embryo, foetus etc. and to manipulate the sex of progeny.

XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment\_scores:

Quality: 34.00 Length: 34

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AAr42757 ..

Align seg 1/1 to: AAr42757 from: 1 to: 153

```
8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLy 24
|||||
52 AAGAAACCCGCGCAAGACCTACAGACCCGTCGACGCTACCTCATTTAA 101
|||||
24 sLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGlyA 41
|||||
102 ATTACTGAAATTAAGACCTTGCAAGAAAGACACAAAGGCTTGCTGTG 151
|||||
41 la 41
||
152 CA 153

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758

seq_documentation_block:
ID AA142758 standard; DNA; 153 BP.
XX
AC AA142758:
XX
XX 12-MAR-1997 (first entry)
XX
DE Chick CHD-W gene fragment.
XX
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;
XX CHD-1A; CHD-W; W chromosome; ss.
XX
OS Gallus sp.
XX
XX Key Location/Qualifiers
XX misc_difference 52..81
XX /*tag= a
XX /*note= "bases 52-81 are a repeat of bases 22-51
XX and are ignored in the translated amino
XX acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX P-PSDB; AAM08148.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAM08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,
XX foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
us-08-973-363-7 x AA142758 ..
Align seg 1/1 to: AA142758 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLy 24
|||||
52 AAGAAACCCGCGCAAGACCTTACAGACCCGTCGACGCTACCTCATTTAA 101
|||||
24 sLeuLeuAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGlyA 41
|||||
102 ATTACTGAAATTAAGACCTTGCAAGAAAGACACAAAGGCTTGCTGTG 151
|||||
41 la 41
||
152 CA 153

seq_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759

seq_documentation_block:
ID AA142759 standard; DNA; 153 BP.
XX
AC AA142759:
XX
XX 12-MAR-1997 (first entry)
XX
DE Great tit CHD-W gene fragment.
XX
XX Bird; sex determination; chromodomain-Helicase-DNA binding 1;
XX CHD-1A; CHD-W; W chromosome; ss.
XX
OS Parus major.
XX
XX Key Location/Qualifiers
XX misc_difference 52..81
XX /*tag= a
XX /*note= "bases 52-81 are a repeat of bases 22-51
XX and are ignored in the translated amino
XX acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX P-PSDB; AAM08149.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A = Avian) gene (AA142757),
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAM08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,
XX foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

Quality: 27.00 Length: 27  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AAT42759 ..

Align seg 1/1 to: AAT42759 from: 1 to: 153

```

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLeu 24
|||||
52 AAGAAACCAAGGCAAGGAGCTTGCAGAGCCCTGCAGATTACTCATTTAA 101
|||||
24 sLeuLeuAsnLysAspLeuAlaArgLysGlu 34
|||||
102 ATTACTGATATAAGACCTTGCAGAAAGAAAGAA 132

```

seq\_name: /SIDSL/gcdata/hold-geneseq/gene-seq-n-emb1/NA1998.DAT:AAV59280

seq\_documentation\_block:

ID AAV59280 standard; cDNA; 1311 BP.

AC AAV59280;

DT 14-DEC-1998 (first entry)

DE Altered telomere repeat binding factor 1 gene.

XX ds; human: telomere repeat binding factor; A-TRF: dimerisation domain;

KW telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

OS Homo sapiens.

OS Synthetic.

FH Key Location/Qualifiers

FT CDS 1..1311

FT /\*tag= a

FT /product= "A-TRF"

XX MO9836066-A1.

PD 20-AUG-1998.

PF 13-FEB-1998; 98WO-US02765.

PR 04-FEB-1998; 98US-0018628.

PR 13-FEB-1997; 97US-0800264.

XX (UTRQ ) UNIV ROCKEFELLER.

PI Bianchi A, De Lange T, Van Steensel B;

XX WPI: 1998-480769/41.

DR P-PSDB: AAM59280.

PT Nucleic acid encoding altered telomere repeat binding protein and  
related vectors - transformants, hetero-dimers and antibodies, used  
to inhibit shortening of telomeres caused by ageing or disease,  
PT also used to extend life of cells in culture

XX Claim 14; Page 110-111; 163pp; English.

PS The altered vertebrate telomere repeat binding protein (A-TRF) has a  
CC telomere repeat binding factor (TRF) dimerisation domain, and forms a  
CC hetero-dimer with TRF, preventing it from binding to the specified repeat  
CC sequence. A-TRF, optionally expressed by gene therapy, is used to  
CC inhibit shortening of telomeres associated with ageing (for cosmetic  
CC purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
CC atrophy of the skin, age-related macular degeneration, atherosclerosis,  
CC tumours and viral (including human immune deficiency virus) infection.  
CC Cells expressing A-TRF also have an increased life span in vitro, e.g.  
CC for expression of recombinant proteins or where intended for subsequent  
CC transplant or for testing, eliminating the need for transformation.

XX S0 Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

alignment\_scores:

Quality: 25.00 Length: 25

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AAV59280 ..

Align seg 1/1 to: AAV59280 from: 1 to: 1311

```

1 IleLeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnThr 17
|||||
1237 ATTCTTCGAGATGATCCTGTATATAAACCAAGCAAGCAAGCAAGTTCAGAGC 1286
|||||
17 rArgAlaAspTyrLeuIleLysLeu 25
|||||
1287 CCGTGCAGACCTACCTCATCAACTA 1311

```

seq\_name: /SIDSL/gcdata/hold-geneseq/gene-seq-n-emb1/NA1997.DAT:AAT42756

seq\_documentation\_block:

ID AAT42756 standard; DNA; 153 BP.

AC AAT42756;

DT 12-MAR-1997 (first entry)

DE Mouse CHD-1 gene (bases 3855-977).

XX Bird; sex determination; chromodomain-helicase-DNA binding 1;

KW CHD-1; CHD-W; W chromosome; ss.

OS Mus sp.

FH Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51

FT and are ignored in the translated amino

FT acid sequence given in Fig 3"

XX MO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

XX (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

XX WPI: 1997-043127/04.

DR P-PSDB: AAM08146.

PT Avian chromodomain-helicase-DNA binding genes determine sex in

PT birds - used for sex determin. and to control sex of progeny

XX Claim 8; Fig 3; 76pp; English.

PS Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
CC to portions of the chicken CHD-1A (A = Avian) gene (see also  
CC and AAT42757), chicken CHD-W (W refers to the W chromosome) gene  
CC (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
CC Translated amino acid sequences of this region are provided in  
CC AAM08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
CC AAT42754-55) genes determine sex in birds and can be used to identify  
CC the sex of an embryo, foetus etc. and to manipulate the sex of

CC progeny.  
XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;  
SQ

alignment\_scores:  
Quality: 19.00 Length: 19  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x AAT42756 ..

Align seg 1/1 to: AAT42756 from: 1 to: 153

8 LysElysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAAAAACCAAGCAAAACAGTTACAGACCGGTGACAGACTCATCA 101  
|||||  
24 sLeuLeu 26  
|||||  
102 ACTACTT 108

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06443

seq\_documentation\_block:  
ID ABL06443 standard; cDNA: 6240 BP.  
XX  
AC ABL06443;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13611.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EM;  
XX  
DR WPI: 2001-656860/75.  
DR P-PSDB; ABB62340.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1: SEQ ID NO 13611; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;  
SQ

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19  
|||||  
4205 AAGCCCGAGCCAGCAGCTGACAGCCGCGCC 4237  
|||||

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06442

seq\_documentation\_block:  
ID ABL06442 standard; cDNA: 9933 BP.  
XX  
AC ABL06442;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13608.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EM;  
XX  
DR WPI: 2001-656860/75.  
DR P-PSDB; ABB62339.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Claim 1: SEQ ID NO 13608; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;  
SQ

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment block:  
US-08-973-363-7 x ABL06442 ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933

9 LysProGlnAlaLysGlnLeuGlnThrGala 19  
|||||  
6745 AACGCCAGCGCCAGCAGCTGCGACGCGTCCC 6777

seq\_name: /SIDSI/gcsgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK88882

seq\_documentation\_block:

ID AAK88882 standard: cDNA: 421 BP.

XX AAK88882;

DT 05-NOV-2001 (first entry)

DE Human digestive system antigen coding sequence SEQ ID NO: 1198.

XX Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ss.

OS Homo sapiens.

PN WO200155314-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01324.

XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.



PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
DR WPI; 2001-457727/49.  
DR P-PSDB; AAM38625.  
XX  
PT Isolated polypeptide for treating, preventing and/or prognosing  
PT disorders related to the colon and rectum including colorectal cancers  
PT and also for testing and detection e.g. diagnosis -  
XX  
XX  
PS Claim 1; SEQ ID NO: 67; 522pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of colorectal cancer antigens. These are shown in  
CC AAI57547-AAI57619 and AAM38569-AAM38641. These can be used in the  
CC diagnosis, prevention and treatment of cancer of the colon and/or rectum.  
CC The present sequence is a colorectal cancer antigen coding sequence of  
CC the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other:  
  
alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-7 x AAI57603 ..  
  
Align seg 1/1 to: AAI57603 from: 1 to: 421  
  
13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
|||||  
146 AAGCAGCTACAGACCGACGCGATCTACTTG 175  
  
seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NAI1998.DAT:AAV33521  
seq\_documentation\_block:  
ID AAV33521 standard; cDNA; 252 BP.  
XX  
XX AAV33521;  
AC  
XX  
XX 29-DEC-1998 (first entry)  
DT  
XX  
XX  
DE Clone 23789 cDNA fragment encoding a human protein kinase homolog.  
XX  
XX  
KW Protein kinase; cell signalling; inflammation; carcinoma; diabetes;  
KW human x-linked agammaglobulinemia; nonpheryocytic haemolytic anaemia;  
KW arteriosclerosis; glioma; restenosis; cholera-based septic shock;  
KW CHKFRNK chicken tyr kinase; ss.  
XX  
OS Homo sapiens.  
XX  
XX US5817479-A.  
PN  
XX  
XX 06-OCT-1998.  
PD  
XX  
XX 07-AUG-1996; 96US-0700575.  
PF  
XX  
XX 07-AUG-1996; 96US-0700575.  
PR  
PA (INCY-) INCYTE PHARM INC.  
XX  
XX  
PI Au-Young J, Bandman O, Hawkins PR, Wilde CG;  
XX  
DR WPI; 1998-556387/47.



```
XX Human kinase poly(nucleotide(s) and recombinant products - useful
PT for identification of modulators of the enzyme, and treatment of
PT diseases associated with abnormal kinase expression
XX
XX Claim 1: Columns 29-30: 30pp; English.
XX
CC The invention provides polynucleotides which encode novel protein
CC kinase homologs expressed in various human cells and tissues. The
CC present sequence represents the clone 23789 cDNA fragment derived
CC from a human inflamed adenoid cDNA library. The cDNA encodes a
CC protein kinase which shows homology to the CKRFRK chicken tyr kinase.
CC Vectors and host cells can be used for recombinant production of the
CC protein kinase homolog. The recombinant proteins may be used to raise
CC antibodies for use as anti-kinase therapeutics. Oligonucleotides based
CC on the polynucleotide sequences, i.e. probes and antisense constructs,
CC the peptides and antibodies are claimed to be useful as tools for
CC studying signaling cascades in cells and proteins, and for identifying
CC inhibitors (drugs) to treat diseases and inflammatory conditions
CC associated with abnormal kinase expression. Diseases that are claimed
CC to be treatable include human X-linked agammaglobulinemia,
CC nonspherocytic hemolytic anaemia, arteriosclerosis, carcinomas,
CC diabetes, glomerulonephritis, restenosis, cholera-based septic shock, etc.
XX
SO Sequence 252 BP; 63 A; 70 C; 72 G; 47 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x AAV33521 ..
Align seg 1/1 to: AAV33521 from: 1 to: 252

24 TytleuLeuAaNLySAsPLeua1a 31
|||||
100 AACTGCTCAACAAAGACCTGCA 123

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA61903

seq_documentation_block:
ID ABA61903 standard; DNA: 544 BP.
XX
XX ABA61903:
XX
XX 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #10208.
XX
XX Human; foetal liver; gene expression: single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
```

```
XX WPI: 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver.
XX
XX Claim 1: SEQ ID NO 10208; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SO Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x ABA61903 ..
Align seg 1/1 to: ABA61903 from: 1 to: 544

21 TytleuLeuAaNLySAsNLyS 28
|||||
243 TACCTGATTAAGTATTAAATAG 266

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK10214

seq_documentation_block:
ID AAK10214 standard; DNA: 544 BP.
XX
XX AAK10214:
XX
XX 05-NOV-2001 (first entry)
XX
XX
DE Human brain expressed single exon probe SEQ ID NO: 10205.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer; ss.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PI
```

PT brains -  
XX  
PS Example 4; SEQ ID NO: 10205; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-7 x AAK10214 ..  
  
Align seg 1/1 to: AAK10214 from: 1 to: 544  
  
21 TyrleuilelystleuLeuAsnLys 28  
|||||  
243 TACCTGATAAGTTATTAATAAG 266  
  
seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:AAK36112  
  
seq\_documentation\_block:  
ID AAK36112 standard; DNA; 544 BP.  
XX  
AC AAK36112;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 10669.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI: 2001-488900/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human bone marrow -  
XX  
PS Example 4; SEQ ID NO: 10669; 658pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.  
XX  
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-7 x AAK36112 ..  
  
Align seg 1/1 to: AAK36112 from: 1 to: 544  
  
21 TyrleuilelystleuLeuAsnLys 28  
|||||  
243 TACCTGATAAGTTATTAATAAG 266

Mon Aug 5 11:51:58 2002

us-08-973-363-7.01ip2n.rng

Page 11



**THIS PAGE BLANK (USPTO)**



```

DEFINITION AL644594 XGC-egg Silurana tropicalis cDNA clone L1E1d12 5', mRNA
sequence.
ACCESSION AL644594
VERSION AL644594.1 GI:16796719
KEYWORDS EST
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 645)
AUTHORS Huckie,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2001 (10_2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Huckie E
Sanger Centre
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: L1E1d12.pic
Sequencing primer: PIC
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
FEATURES
source
location/Qualifiers
1..645
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="L1E1d12"
/clone_lib="XGC-egg"
/dev_stage="egg"
/lab_host="Escherichia coli XL1-blue"
/note="vector: PCS107; site_1: EcoRI; site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into PCS107 with
EcoRI at the 5' end and NotI at the 3' end"
BASE COUNT 222 a 125 c 156 g 141 t 1 others
ORIGIN

alignment_scores:
Quality: 30.00 Length: 30
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x AL644594 ..
Align seg 1/1 to: AL644594 from: 1 to: 645

1 l1eleuproaspappproasp1yslpsproglalalysglnleuglnth 17
|||||
456 ATTTCACGATGATCCAGATAGAGAGCCCAAGCAGCAGCTACAGAC 505
|||||
17 rargalasp1tyrleuileuileu1e1ysleu1e1aspl1eu 30
|||||
506 CAGAGCTGACTACCTCATTAACCTTCATTAAGATCTG 545
|||||
seq_name: gb_est1:AL601246

seq_documentation_block:
LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001
DEFINITION DKFZP313J1040_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone
DKFZP313J1040 5', mRNA sequence.
ACCESSION AL601246
VERSION AL601246.1 GI:15164752
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 430)
AUTHORS Blocker,H., Boecher,M., Brandt,P., Mewes,W., Weil,B. and Wiemann
,S.

```

```

TITLE EST (Blocker,H., Boecher,M., Brandt,P., Mewes,H.W., Weil,B. and
Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Blocker H
MIPS
Am Klopfersplitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZP313J1040) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY. Email: clone@rzpd.de.
FEATURES
source
location/Qualifiers
1..430
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP313J1040"
/clone_lib="313 (synonym: hlcc2)"
/dev_stage="adult"
/lab_host="DH10B"
/note="vector: pT7Blue; site_1: SfiI; site_2: SfiI;
cDNA-collection"
BASE COUNT 163 a 81 c 86 g 100 t
ORIGIN

alignment_scores:
Quality: 26.00 Length: 26
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-7 x AL601246 ..
Align seg 1/1 to: AL601246 from: 1 to: 430

1 l1eleuproaspappproasp1yslpsproglalalysglnleuglnth 17
|||||
218 ATTCTTCAGATGATCCGATTAAGAGACCAAGCAAGAGTTGACGAC 267
|||||
17 rargalasp1tyrleuileuileu1e1ysleu1e1aspl1eu 26
|||||
268 CCGTCGAGACTACCTCATTAATTAATT 295
|||||
seq_name: gb_est1:A1890775

seq_documentation_block:
LOCUS A1890775 547 bp mRNA linear EST 07-MAR-2000
DEFINITION wmg5f11.x1 NCI_CGAP_U22 Homo sapiens cDNA clone IMAGE:2443725 3',
similar to SW:CHD1_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING
PROTEIN 1 ; , mRNA sequence.
ACCESSION A1890775
VERSION A1890775.1 GI:5595939
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 547)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@b3.rnmail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

```



1 HleuProaspPProaspPlysProglAlaLysGlnleuGlnh 17  
 |||TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT|  
 489 ATCTTCCAGATGATCCTGATTAATAAACCAAGCAAAAGATTACAGAC 538  
 17 rArgAlaAspTyrLeuIleLysLeuLeu 26  
 |||TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT|  
 539 CCGTGCAGACTACTCTCATCAAACTACTT 566

seq\_name: gb\_est1:BB461065

seq\_documentation\_block:  
 LOCUS BB461065 660 bp mRNA linear EST 25-OCT-2001  
 DEFINITION BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion  
 Mus musculus cDNA clone D130070B13 3' similar to I10410 Mouse  
 DNA-binding protein (CHD-1) mRNA, mRNA sequence.  
 BB461065  
 BB461065.2 GI:16426612  
 EST.  
 house mouse.  
 Mus musculus  
 ORGANISM

REFERENCE  
 AUTHORS  
 1 (bases 1 to 660)  
 Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Koda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
 Unpublished (2001)  
 On Jul 21, 2000 this sequence version replaced gi:9356558.  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsr.riken.go.jp.  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci,P., Shibata,Y., Hayatsu,M., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 Wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for  
 further details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.

FEATURES  
 Source 1. 660  
 Location/Qualifiers

/organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="D130070B13"  
 /clone\_lib="RIKEN full-length enriched, 12 days embryo  
 spinal ganglion"  
 /tissue\_type="spinal ganglion"  
 /dev\_stage="12 days embryo"  
 /lab\_host="DH10B"  
 /note="Site 1: Sall1; Site 2: BamHI; cDNA library was  
 prepared and sequenced in Mouse Genome Encyclopedia  
 Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in  
 RIKEN. Division of Experimental Animal Research in Riken  
 contributed to prepare mouse tissues. 1st strand cDNA was  
 primed with a primer [5',  
 GAGAGAGAGCGCGGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was  
 prepared by using trehalose thermo-activated reverse  
 transcriptase and subsequently enriched for full-length by  
 cap-trapper. Second strand cDNA was prepared with the  
 primer adapter of sequence [5',  
 GAGAGAGAGATTCTCGAGTTTAAATTAAATCCCGCCCCCCC 3']. cDNA  
 was cleaved with BamHI and XhoI. Vector: a modified  
 pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 211 a 130 c 168 g 151 t

ORIGIN

alignment\_scores:  
 Quality: 26.00 Length: 26  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-7 x BB461065 ..

Align seg 1/1 to: BB461065 from: 1 to: 660

1 HleuProaspPProaspPlysProglAlaLysGlnleuGlnh 17  
 |||TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT|  
 481 ATCTTCCAGATGATCCTGATTAATAAACCAAGCAAAAGATTACAGAC 530  
 17 rArgAlaAspTyrLeuIleLysLeuLeu 26  
 |||TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT|  
 531 CCGTGCAGACTACTCTCATCAAACTACTT 558

seq\_name: gb\_est1:AUI25712

seq\_documentation\_block:  
 LOCUS AUI25712 866 bp mRNA linear EST 23-OCT-2000  
 DEFINITION AUI25712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
 sequence.  
 AUI25712  
 AUI25712.1 GI:10950428  
 EST.  
 HRI human  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 866)  
 Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
 Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
 Isogai,T.  
 HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
 Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
 Y., Sugano,S., Isogai,T.)  
 Unpublished (2000)  
 Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3951  
 Fax: 81-438-52-3952  
 Email: genomics@hri.co.jp

JOURNAL  
 COMMENT



HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

## FEATURES

source

1. 866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="NT2RM4002061"  
/clone\_lib="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

## alignment\_scores:

Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

1 HleuPProaspPaspProaspPlyslsProglAlalysGlnLeuGlnth 17  
|||||  
450 ATTCTCCAGATGATCCGATTAATAAACACAGCAAGCAAGAGTGGACGAC 499  
|||||  
17 rArgAlaSPfYrLeuIlleLysLeuLeu 26  
|||||  
500 CCGTCGACGACTACCTCATCAATTAATCTT 527

seq\_name: gb\_estc:BE895133

## seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION B014360801 NIH\_MGC\_72 Homo sapiens cDNA IMAGE:3921087 5',  
mRNA sequence.

ACCESSION BE895133  
VERSION BE895133.1 GI:10358221

KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 1028)

NIH-MGC <http://mgi.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: ATCC/DCTP/DRP

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Place: L14M9753 row: h column: 16

High quality sequence stop: 488.

Location/Qualifiers

1. 1028

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_id="NIH\_MGC\_72"

/clone\_lib="NIH\_MGC\_72"

/tissue\_type="melanotic melanoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: skin; Vector: PCMV-SPORT6; Site\_1: NotI;

site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life Technologies."

## FEATURES

BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

## alignment\_scores:

Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-7 x BE895133 ..

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 HleuPProaspPaspProaspPlyslsProglAlalysGlnLeuGlnth 17  
|||||  
131 ATTCTCCAGATGATCCGATTAATAAACACAGCAAGCAAGAGTGGACGAC 180  
|||||  
17 rArgAlaSPfYrLeuIlleLysLeuLeu 26  
|||||  
181 CCGTCGACGACTACCTCATCAATTAATCTT 208

seq\_name: gb\_estl:AW96787

## seq\_documentation\_block:

LOCUS AW96787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW96787  
VERSION AW96787.1 GI:8257021

KEYWORDS EST.  
SOURCE human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

## REFERENCE

AUTHORS

1 (bases 1 to 337)

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagal,M.A., de Silva,W.J.R., Zago,M.A., Bordin,S., Costa,F.F.,  
Gadgil,G.H., Carvalha,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J., and  
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

## JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=8t2-QV3-BN0047-230>

200-102-d03&t3=2000-02-23&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 2

High quality sequence stop: 337.

Location/Qualifiers

1. 337

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_id="BN0047"

/dev\_stage="Adult"

/note="Organ: breast-normal; Vector: puc18; Site\_1: SmaI;

Site\_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

ORIGIN

alignment\_scores:                   Quality: 23.00                   Length: 23  
                                  Ratio: 1.000                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x AW996787/rev ..

Align seg 1/1 to reverse of: AW996787 from: 1 to: 337

```
1 1tleuProaspProaspLysProglAlaLysGlnLeuGlnth 17
|||||
105 ATTCTTCAGATGATGCCGATAAAAACCAAGCAAGCAAGTTGACAGAC 56
17 rargAlaasPTyrlenu 23
|||||
55 CCGTGCAGACTATCTCATC 37
```

seq\_name: gb\_est1:BB830730

seq\_documentation\_block:

LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527  
Jyg-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA  
sequence.

ACCESSION BB830730  
VERSION BB830730.1 GI:17008973

KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)  
CONTACT: Yoshihide Hayashizaki

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carlini,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
wael,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
Location/Qualifiers  
source 1. .438  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930013K04"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jyg-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jyg-MC(B)"

BASE COUNT 138 a 96 c 108 g 96 t

ORIGIN

alignment\_scores:                   Quality: 22.00                   Length: 22  
                                  Ratio: 1.000                   Gaps: 0  
Percent Similarity: 100.000       Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x BB830730 ..

Align seg 1/1 to: BB830730 from: 1 to: 438

```
1 1tleuProaspProaspLysProglAlaLysGlnLeuGlnth 17
|||||
300 ATTCTTCAGATGATGCCGATAAAAACCAAGCAAGCAAGTTACAGAC 349
17 rargAlaasPTyrlenu 22
|||||
350 CCGTGCAGACTATCTCATC 365
```

seq\_name: gb\_est1:BB834922

seq\_documentation\_block:

LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
Jyg-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA  
sequence.

ACCESSION BB834922  
VERSION BB834922.1 GI:17013165

KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Wataniki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)  
CONTACT: Yoshihide Hayashizaki

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
Wagl, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Mitalik, K., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)  
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
e mouse tissues.

#### FEATURES

source  
Location/Qualifiers  
1. 446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone\_lib="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jy9-MC(B)"  
BASE COUNT 139 a 99 c 108 g 100 t  
ORIGIN

#### alignment\_scores:

Quality: 22.00 Length: 22  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-08-973-363-7 x BB834922 ..  
Align seg 1/1 to: BB834922 from: 1 to: 446

1 11leuPpOaSpAsPrOaSpLySlySProGlnAlaLysGlnLeuGlnth 17  
|||||  
309 ATTCTTCCAGATGATCTCATATAAAACACAGCCAAACAGTTACGAC 358  
17 rArgAlaAspTyrLeu 22  
|||||  
359 CCGTGCACACTACTCTC 374

seq\_name: gb\_est2:BF239967

seq\_documentation\_block:

LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000  
DEFINITION M01905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
mRNA sequence.

ACCESSION BF239967  
VERSION BF239967.1 GI:11153890  
KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE NIH-MGC <http://mgc.ncl.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: [cgabbs+email.nih.gov](mailto:cgabbs+email.nih.gov)

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Plate: LNCMI033 row: k column: 18  
High quality sequence stop: 562.  
Location/Qualifiers

#### FEATURES

source  
1. 821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_54"  
/tissue\_type="from chronic myelogenous leukemia"  
/lab\_host="DH10B (11 phage-resistant)"  
/note="Organ: bone marrow; Vector: pDNK-LIB (Clontech); Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggcctatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCGCGGCGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."  
BASE COUNT 345 a 149 c 190 g 137 t  
ORIGIN

#### alignment\_scores:

Quality: 21.00 Length: 21  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-08-973-363-7 x BF239967 ..  
Align seg 1/1 to: BF239967 from: 1 to: 821

6 ProAspLySlySProGlnAlaLysGlnLeuGlnThrArgAlaSpTyrLe 22  
|||||  
3 CCGGATTAATAAACACCAACAAACAGTTGACAGCCGTGCAGACTACT 52  
22 uileLysLeuLeu 26  
|||||  
53 CATCAATTAATCTT 65

seq\_name: gb\_est1:AW97058

seq\_documentation\_block:

LOCUS AW97058 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BR0047-150400-152-C03 BR0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW97058  
VERSION AW97058.1 GI:8257292  
KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 686)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

Nagai, M.A., de Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.,

Goldman, G.H., Garvalho, A.F., Matsukuma, A., Beira, G.S., Simpson, D.H.,

Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare

, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and

Simpson, A.V.  
Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL Contact: Simpson A.J.G.

COMMENT Laboratory of Cancer Genetics

Medline Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=ft2-QV3-BN0047-150>)  
 400-152-c03&t3=2000-04-15&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 19  
 High quality sequence stop: 678.  
 Location/Qualifiers

## FEATURES

source

1..686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BN0047"  
 /dev\_stage="Adult"  
 /note="Organ: breast\_normal; Vector: puc18; Site\_1: Sma1; Site\_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 165 a 154 c 126 g 241 t  
 ORIGIN

## alignment\_scores:

Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-7 x AW997058/rev ..

Align seg 1/1 to reverse of: AW997058 from: 1 to: 686

9 LysProGlnAlaIysGlnLeuGlnThrArgAlaAspTyrLeuIleIysLe 25  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 104 AAACCCACAGCAAAACAGTTGCAGACCCGTGCAGACTCACTCAATTT 55

25 ulau 26  
 ||||

54 ACTT 51

seq\_name: gb\_gss:CNS04DVC

## seq\_documentation\_block:

LOCUS CNS04DVC 856 bp DNA linear GSS 21-MAY-2000  
 DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone  
 103K08 of library G from Tetradon nigroviridis, genomic survey  
 sequence.  
 ACCESSION AL286261  
 VERSION AL286261.1 GI:8024707  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetradon nigroviridis.  
 ORGANISM Tetradon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 Tetraodontidae; Tetradon.  
 1 (bases 1 to 856)  
 Roest-Crollius,H., Jallion,O., Dasilva,C., Fitzames,C., Fisher,C.,  
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
 Weissenbach,J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetradon nigroviridis  
 Unpublished  
 2 (bases 1 to 856)  
 Roest-Crollius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,  
 Bernot,A., Fitzames,C., Wincker,P., Brotier,P., Quetier,F.,

REFERENCE  
 AUTHORS

## TITLE

JOURNAL  
 REFERENCE  
 AUTHORS

Saurin,W. and Weissenbach,J.  
 Human gene number estimate provided by genome wide analysis using  
 Tetradon nigroviridis DNA sequence  
 Unpublished  
 3 (bases 1 to 856)  
 Genoscope.  
 Direct Submission  
 Submitted (12-Apr-2000) to the EMBL/Genbank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetradon nigroviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetradon>.  
 Location/Qualifiers

## FEATURES

source

1..856  
 /organism="Tetradon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone="103K08"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0BG103BF04LP1-end : T7"  
 BASE COUNT 84 a 308 c 252 g 206 t 6 others  
 ORIGIN

## alignment\_scores:

Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-7 x CNS04DVC/rev ..

Align seg 1/1 to reverse of: CNS04DVC from: 1 to: 856

2 LeuProAspAspProAspLysLysProGlnAlaIysGlnLeuGlnThrArg 18  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 674 CTGCCGATGATCTCTATTAAGAAGCTCAGGCCAAGCATTACAGACAG 625

18 gAla 19  
 ||||

624 AGCC 621

seq\_name: gb\_est2:BF881342

## seq\_documentation\_block:

LOCUS BF881342 129 bp mRNA linear EST 17-JAN-2001  
 DEFINITION PMO-ET0208-031200-001-f12 ET0208 Homo sapiens CDNA, mRNA sequence.  
 ACCESSION BF881342  
 VERSION BF881342.1 GI:12271468  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 AUTHORS

1 (bases 1 to 129)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.G.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome

JOURNAL  
 MEDLINE  
 COMMENT

Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM0&l2=PM0-ET0208-031200-001-f12&l3=2000-12-03&l4=1)

Seq primer: puc 18 forward  
High quality sequence start: 22  
High quality sequence stop: 128.

# FEATURES

source

1. 129  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="ET0208"  
/dev\_stage="Adult"  
/note="Organ: lung tumor; Vector: puc18; site: 1; Smal:  
Site 2: Smal: A mini-library was made by cloning products  
derived from ORSRES PCR (U S letters patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 21 a 48 c 23 g 37 t  
ORIGIN.

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-7 x BF881342/rev ..

Align seg 1/1 to reverse of: BF881342 from: 1 to: 129

13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
|||||  
108 AAGCAGCTACGACCGAGCGATTACTTG 79

**THIS PAGE BLANK (USPTO)**



```

COMMENT      Other publication AV 5906996 961224.
FEATURES
  source      Location/Qualifiers
              1..6608
              /organism="unidentified"
              /db_xref="taxon:32644"
BASE COUNT   2289 a 1207 c 1459 g 1653 t
ORIGIN
alignment_scores:
  Quality:    41.00      Length:    41
  Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-08-973-363-8 x AS58691 ..
Align seg 1/1 to: AS58691 from: 1 to: 6608
1 11eLeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17
  |||||
4080 ATTATACCTGATGATCCACAGACAAGAAACCCAGGCAAGAGCTACAGAC 4129
17 rArgAlaAspTyrLeuIleLysLeuAsnLysAspLeuAlaArgLysG 34
  |||||
4130 CCGTGCAGACTACCTCATTTAAATTACTGATTAAGACCTTGCAAGAAAG 4179
34 1uAlaGlnArgLeuAlaGlyAla 41
  |||||
4180 AAGCACAAAGCGCTTGCTGCTGCCA 4202
seq_name: gb_ov:AF004397
seq_documentation_block:
LOCUS      AF004397      6872 bp      mRNA      linear      VRT 08-OCT-1997
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the Z chromosome
            protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete
            cds.
ACCESSION   AF004397
VERSION     AF004397.1  GI:2501845
KEYWORDS
SOURCE      chicken.
ORGANISM    Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 6872)
            Griffiths,R. and Korn,R.M.
            A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus
            JOURNAL Gene 197 (1-2), 225-229 (1997)
            MEDLINE 97473516
REFERENCE   2 (bases 1 to 6872)
            Griffiths,R. and Korn,R.M.
            Direct Submission
            Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,
            Glasgow G12 8QQ, UK
FEATURES
  source      Location/Qualifiers
              1..6872
              /organism="Gallus gallus"
              /db_xref="taxon:9031"
              1..6872
              /gene="CHD-Z"
              228..5654
              /gene="CHD-Z"
              /function="role in chromatin architecture"
              /note="CHD protein with hydrophilic domain"
              /codon_start=1
              /product="chromo-helicase-DNA-binding on the Z chromosome
              protein"
              /protein_id="AAC60282.1"
              /db_xref="GI:2501845"
              /translation="MNGHSDSEYVRNSGESSRDDDSGSAGSGSSGSSSDGSS
              SOSGSSSESGSGSSESDTSREKKQVQAPPRKADGEFWRKSPSIIAYVRSAV

```

[illegible]



```

BIRDS
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;
COMMENT ISIS INNOVATION (GB)
FEATURES Other publication AU 5906996 961224.
SOURCE Location/Qualifiers
1.153
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 58 a 40 c 31 g 24 t
ORIGIN

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-8 x A58684 ..
Align seg 1/1 to: A58684 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLy 24
|||||
52 AAGAAACCCAGCGCAAGCAGCTACAGCCGCGAGACTACCTCATTTAA 101
|||||
102 ATTACTGAATTAAGACCTTCAGAAAGAAAGCAAGCCTGCTGTG 151
|||||
41 1a 41
152 CA 153

seq_name: gb_pat:A58685
seq_documentation_block:
LOCUS A58685 153 bp DNA linear PAT 06-MAR-1996
DEFINITION Sequence 4 from Patent WO9639505.
ACCESSION A58685
VERSION A58685.1 GI:3714248
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 153)
AUTHORS Griffiths R. and Tlwardt B.
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL Patent: WO 9639505-A 4 12-DEC-1996;
COMMENT ISIS INNOVATION (GB)
FEATURES Other publication AU 5906996 961224.
SOURCE Location/Qualifiers
1.153
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 56 a 36 c 31 g 30 t
ORIGIN

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-8 x A58685 ..
Align seg 1/1 to: A58685 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLy 24
|||||
52 AAGAAACCCAGCGCAAGCAGCTACAGCCGCGAGACTACCTCATTTAA 101
|||||
102 ATTACTGAATTAAGACCTTCAGAAAGAAAGCAAGCCTGCTGTG 151
|||||
41 1a 41
152 CA 153

seq_name: gb_pat:A58686
seq_documentation_block:
LOCUS A58686 153 bp DNA linear PAT 06-MAR-1996
DEFINITION Sequence 5 from Patent WO9639505.
ACCESSION A58686
VERSION A58686.1 GI:3714249
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 153)
AUTHORS Griffiths R. and Tlwardt B.
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS
JOURNAL Patent: WO 9639505-A 5 12-DEC-1996;
COMMENT ISIS INNOVATION (GB)
FEATURES Other publication AU 5906996 961224.
SOURCE Location/Qualifiers
1.153
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 60 a 34 c 31 g 28 t
ORIGIN

alignment_scores:
Quality: 27.00 Length: 27
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-8 x A58686 ..
Align seg 1/1 to: A58686 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLy 24
|||||
52 AAGAAACCCAGCGCAAGCAGCTTCAGAAAGAAAGCAAGCCTGCTGTG 101
|||||
102 ATTACTGAATTAAGACCTTCAGAAAGAAAGCAAGCCTGCTGTG 151
|||||
41 1a 41
152 CA 153

seq_name: gb_ro:MUSCHD1X
seq_documentation_block:
LOCUS MUSCHD1X 5349 bp mRNA linear ROD 19-SEP-1996
DEFINITION Mouse DNA-binding protein (CHD-1) mRNA, complete cds.
ACCESSION U10410 X66028
VERSION U10410.1 GI:455014
KEYWORDS DNA binding protein.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 5349)
AUTHORS Delmas Y., Stokes D.G. and Perry R.P.
TITLE A mammalian DNA-binding protein that contains a chromodomain and an POU/Natl. Acad. Sci. U.S.A. 90 (6), 2414-2418 (1993)
JOURNAL MEDLINE 93211972
MEDLINE 2 (bases 1 to 5349)
Perry, R.P.

```

TITLE Direct Submission  
JOURNAL Submitted (08-APR-1993) Robert Perry, The Fox Chase Cancer Center,  
7701 Burholme Avenue, Philadelphia, PA 19111, USA  
COMMENT On Feb 16, 1994 this sequence version replaced gi:293322.  
FEATURES  
Location/Qualifiers

source  
1..5349  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/cell\_type="plasmacytoma"  
1..5349  
/gene="CHD-1"  
171..5306  
/gene="CHD-1"  
/codon\_start=1  
/product="DNA-binding protein"  
/protein\_id="AAB08486.1"  
/db\_xref="GI:455015"  
/translation="MNGHSDSESVNRSGSGSSQSDGDCGSASGSGSSGSSSDGSS  
SQSGSSDSDSGSGSSESDTRENKVAQKPPKVDGAEPKWSPSILAQRASAML  
RKPOQAQOORPASNSGSESDSSSESDSSGAKRKHNDEDMOMSGSPOLG  
SDSESEERDKSSCDGTESDYEPKKNKVRKPNRKSNGKKILGOKKROIDSSE  
DDEDDNDKRSRRQATVNSYKKEDEMKTOSDILEVCGEVDPQPEDEEPTIRRM  
DCRVGRKATGATTTTAYEADGDPNAGPERKKEPGDIOYLTKMGWSHHTWETEE  
TLKOQNVGMRKLDYKRRKQDETKRMNLNASPEDEYVNCQDELTDLHKQIVERI  
IAHSNOKSAAGLPDYCKWQGLPYSECSMEDALISKKEOTCIDYFSNRQKTPFK  
DCKVLKORPREVALKQPSYIGHEGLRDYQLNGMLIAHSCKNSCLADBMGL  
GKTIOITISFLNLYFHEHQLGPELLVPLSTLSMREIOTMSQMANAVVYIGDINSR  
NMIRTHEMHQOTKRLKNILLTYEILLKKAFLGMAFAGYDEAHLKNDLSL  
KYLIDFSNHRLLITGTPLONSLKEMLSLHPEKFSWEDPEEHEGKREGYA  
SLHKELEPLLRLRVKRDVEKSLPAKVEQILRMESALOKQYKWLITRYNKLKSGK  
GSTGFLIMMELKCCNHCYLIPKPDNNEFYNKQDALHILRSSGKLLIDLKILRL  
RERGNRVLIIFSOMVRMLDILAEYLKYRQFPORLDGSELKQALDHFNEGSEDE  
CELSTIRAGGGINLASADTVIIFDSMNPONDLOAARAHRIQOKOVNIYLVTKG  
SVEDEILERRAKKMWLDHLYIORMDTGKTVLHTGSASSSPKPKKEELSLTKGAE  
ELFKPEEBOEPQMDIDEIKRAETHNEPGPLSVGDELLSQRYANFSMDEDDI  
ELEPERKSNMEELIPEQORRLEEERKQKLELELYMLPRMNCAKQISFNSEGRS  
RSRRYSGSDSDSISERRKPRKRGPRRTIPRENKGSDAEIRFIRKSYKKEGPERL  
DAIADALVIRKQESDQNSDONSVATTHVIRNPDMLRLEKNTNHDSDSRYSDRHL  
VOVNAKLIAHDELDLPLHKSIPSDPEERKQYTIICHTKAHFIDMGKSGPTFISG  
IYEVGSGWEMIKMDPDLSTLHKILPDDPKKPOAKOLOTADYLIKLSRLAKREA  
ORLCGAGSKRRKRTAKSKKAMKSTIKVKEIKSDSPLESEKDEDDKLNDSKESK  
DRSKSVSDAIVHITASGEPIVIAESEELQKTFISICKERMRPVKALKOLDPK  
GLSERQLEHTROCLIKIGDHTTECLKEYSNEQIKWRKNLIMFVSKFTEDAKRLH  
KLYKAIRKROESQNSDONSVATTHVIRNPDMLRLEKNTNHDSDSRYSDRHL  
OYHDKHRHOGDSYKSKSRKRPYSFSGNSKDHREMDHYRODSRYYSRPREKRLDD  
HRSRHSNLEGLKDRSHSDHSHDRSHSDHRSSEHTHKKSRDYRYLSDMQLD  
HRAASSGPRPLDORSPYGSRSPEHSHHSTPETHWSSRKT"

BASE COUNT 1739 a 1067 c 1319 g 1224 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x MUSCHDIX ..

Align seg 1/1 to: MUSCHDIX from: 1 to: 5349

1 11leuProaspasProaspIysIysProGlnAlaIatysGlnleuGlnth 17  
|||||  
4026 ATTCTTCCCATGATCTCTGATATAAAAAACCAAGCAAAACAGTTACAGAC 4075  
17 TArgAlaAspTyrlleuIleIysleuLeu 26  
|||||  
4076 CCGTGCAGACTACCTCATCAACTACTT 4103

seq\_name: gb.pr:AF006513  
seq\_documentation\_block:  
LOCUS AF006513 5947 bp mRNA linear PRI 27-NOV-1997

DEFINITION Homo sapiens CHD1 mRNA, complete cds.  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 5947)  
REFERENCE  
Woodage,T., Basrai,M.A., Baxevasis,A.D., Hietter,P. and Collins,F.S.  
Characterization of the CHD family of proteins  
Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)  
MEDLINE  
97470991  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Direct Submission  
Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human  
Genome Research Institute, National Institutes of Health, 49  
Convent Drive, Bethesda, MD 20892-4442, USA  
Location/Qualifiers  
1..5947  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q15-21; near WI-5811"  
1..5947  
/gene="CHD1"  
164..5293  
/gene="CHD1"  
/codon\_start=1  
/product="CHD1"  
/protein\_id="AAB87381.1"  
/db\_xref="GI:2645429"  
/translation="MNGHSDSESVNRSGSGSSQSDGDCGSASGSGSSGSSSDGSS  
SQSGSSDSDSGSGSSESDTRENKVAQKPPKVDGAEPKWSPSILAQRASAML  
RKPOQAQOORPASNSGSESDSSSESDSSGAKRKHNDEDMOMSGSPOLG  
SDSESEERDKSSCDGTESDYEPKKNKVRKPNRKSNGKKILGOKKROIDSSE  
DDEDDNDKRSRRQATVNSYKKEDEMKTOSDILEVCGEVDPQPEDEEPTIRRM  
DCRVGRKATGATTTTAYEADGDPNAGPERKKEPGDIOYLTKMGWSHHTWETEE  
TLKOQNVGMRKLDYKRRKQDETKRMNLNASPEDEYVNCQDELTDLHKQIVERI  
IAHSNOKSAAGLPDYCKWQGLPYSECSMEDALISKKEOTCIDYFSNRQKTPFK  
DCKVLKORPREVALKQPSYIGHEGLRDYQLNGMLIAHSCKNSCLADBMGL  
GKTIOITISFLNLYFHEHQLGPELLVPLSTLSMREIOTMSQMANAVVYIGDINSR  
NMIRTHEMHQOTKRLKNILLTYEILLKKAFLGMAFAGYDEAHLKNDLSL  
KYLIDFSNHRLLITGTPLONSLKEMLSLHPEKFSWEDPEEHEGKREGYA  
SLHKELEPLLRLRVKRDVEKSLPAKVEQILRMESALOKQYKWLITRYNKLKSGK  
GSTGFLIMMELKCCNHCYLIPKPDNNEFYNKQDALHILRSSGKLLIDLKILRL  
RERGNRVLIIFSOMVRMLDILAEYLKYRQFPORLDGSELKQALDHFNEGSEDE  
CELSTIRAGGGINLASADTVIIFDSMNPONDLOAARAHRIQOKOVNIYLVTKG  
SVEDEILERRAKKMWLDHLYIORMDTGKTVLHTGSASSSPKPKKEELSLTKGAE  
ELFKPEEBOEPQMDIDEIKRAETHNEPGPLSVGDELLSQRYANFSMDEDDI  
ELEPERKSNMEELIPEQORRLEEERKQKLELELYMLPRMNCAKQISFNSEGRS  
RSRRYSGSDSDSISERRKPRKRGPRRTIPRENKGSDAEIRFIRKSYKKEGPERL  
DAIADALVIRKQESDQNSDONSVATTHVIRNPDMLRLEKNTNHDSDSRYSDRHL  
VOVNAKLIAHDELDLPLHKSIPSDPEERKQYTIICHTKAHFIDMGKSGPTFISG  
IYEVGSGWEMIKMDPDLSTLHKILPDDPKKPOAKOLOTADYLIKLSRLAKREA  
ORLCGAGSKRRKRTAKSKKAMKSTIKVKEIKSDSPLESEKDEDDKLNDSKESK  
DRSKSVSDAIVHITASGEPIVIAESEELQKTFISICKERMRPVKALKOLDPK  
GLSERQLEHTROCLIKIGDHTTECLKEYSNEQIKWRKNLIMFVSKFTEDAKRLH  
KLYKAIRKROESQNSDONSVATTHVIRNPDMLRLEKNTNHDSDSRYSDRHL  
OYHDKHRHOGDSYKSKSRKRPYSFSGNSKDHREMDHYRODSRYYSRPREKRLDD  
HRSRHSNLEGLKDRSHSDHSHDRSHSDHRSSEHTHKKSRDYRYLSDMQLD  
HRAASSGPRPLDORSPYGSRSPEHSHHSTPETHWSSRKT"

BASE COUNT 2130 a 1004 c 1243 g 1570 t  
ORIGIN  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:

US-08-973-363-8 x AF006513

Align seg 1/1 to: AF006513 from: 1 to: 5947

1 HleleuProaSpAspProaSpLysProGlnAlaLysGlnLeuGlnTh 17  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 4025 ATTCTTCAGATGATCCGATTAACCAACCAACCAACAGTTGCAGAC 4074  
 17 TATGAlaSPtYrLeuIlleLysLeuLeu 26  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 4075 CCGTGCAGACTACCTCATCAATTAATTCTT 4102

seq\_name: gb\_pr:AC092372

seq\_documentation\_block:

LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001  
 DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
 ACCESSION AC092372  
 VERSION AC092372.3 GI:17402768  
 KEYWORDS HTG.

SOURCE

ORGANISM

human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS 1 (bases 1 to 101220)  
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
 JOURNAL Direct Submission

2 (bases 1 to 101220)  
 TITLE DOE Joint Genome Institute.  
 JOURNAL Direct Submission

3 (bases 1 to 101220)  
 TITLE Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

4 (bases 1 to 101220)  
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
 JOURNAL Direct Submission

Submitted (01-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA

On Dec 7, 2001 this sequence version replaced gi:15290448.  
 Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center

www.shgc.stanford.edu  
 Quality: Phrap quality >=40 100% of Sequence;  
 Estimated Total Number of Errors is 0.

NOTE: This insert is not the entire sequence of the clone (entire  
 sequence is 146,7kb). It is clipped at the overlap with AC012624.

The number of bases overlapped is 90404.

Location/Qualifiers

1. 101220  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="RP11-58M12"

BASE COUNT 34122 a 18862 c 17827 g 30409 t

ORIGIN

alignment\_scores:  
 Quality: 26.00 Length: 26  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-8 x AC092372/rev ..

Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220

1 HleleuProaSpAspProaSpLysProGlnAlaLysGlnLeuGlnTh 17  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 26932 ATTCTTCAGATGATCCGATTAACCAACCAACCAACAGTTGCAGAC 26883  
 17 TATGAlaSPtYrLeuIlleLysLeuLeu 26

||||||||||||||||||||||||||||||||||||||||||||||  
 26882 CCGTGCAGACTACCTCATCAATTAATTCTT 26855

seq\_name: gb\_pr:AC012624

seq\_documentation\_block:

LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001  
 DEFINITION Homo sapiens chromosome 5 clone CID-2082117, complete sequence.  
 ACCESSION AC012624  
 VERSION AC012624.6 GI:14993679  
 KEYWORDS HTG.

SOURCE

ORGANISM

human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 134365)  
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
 JOURNAL Direct Submission

2 (bases 1 to 134365)  
 TITLE DOE Joint Genome Institute.  
 JOURNAL Direct Submission

3 (bases 1 to 134365)  
 TITLE Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

4 (bases 1 to 134365)  
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
 JOURNAL Direct Submission

Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA

On Jul 21, 2001 this sequence version replaced gi:14277267.  
 Location/Qualifiers

1. 134365  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CID-2082117"

BASE COUNT 40414 a 24457 c 25503 g 43951 t

ORIGIN

alignment\_scores:  
 Quality: 26.00 Length: 26  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-8 x AC012624 ..

Align seg 1/1 to: AC012624 from: 1 to: 134365

1 HleleuProaSpAspProaSpLysProGlnAlaLysGlnLeuGlnTh 17  
 ||||||||||||||||||||||||||||||||||||||||||||||||  
 118247 ATTCTTCAGATGATCCGATTAACCAACCAACCAACAGTTGCAGAC 118296

17 TATGAlaSPtYrLeuIlleLysLeuLeu 26

118297 CCGTGCAGACTACCTCATCAATTAATTCTT 118324

seq\_name: gb\_htg:AC021449

seq\_documentation\_block:

LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000  
 DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered  
 pieces.  
 ACCESSION AC021449  
 VERSION AC021449.3 GI:10047806  
 KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.

ORGANISM	TITLE	JOURNAL	COMMENT
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
2 (bases 1 to 143079)			
Birtner, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,			
Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,			
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,			
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,			
DeRellano, K., Dewar, K., Domingo, M., Doyle, M., Fenster, J.,			
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,			
Gerday, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,			
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,			
Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, R.,			
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,			
McNeeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,			
Norman, C. H., O'Connor, T., O'Donnell, P., Ollivar, T. M., Peterson, K.,			
Petrie, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rotman, D.,			
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,			
Stojanovic, N., Subramanian, A., Talamas, J., Testfay, S., Theodore, J.,			
Titrell, A., Vassiliev, R., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,			
Zimmer, A. and Zody, M.			
Direct Submission			
Submitted (15-JAN-2000) Whitehead Institute/MIT Center for Genome			
Research, 320 Charles Street, Cambridge, MA 02141, USA			
On Sep 10, 2000 this sequence version replaced by:7407965.			
All repeats were identified using RepeatMasker:			
Smit, A. F. A. & Green, P. (1996-1997)			
http://ftp.genome.washington.edu/RM/RepeatMasker.html			
Center: Whitehead Institute/ MIT Center for Genome Research			
Center code: WIBR			
Web site: http://www-seq.wi.mit.edu			
Contact: sequence_submissions@genome.wi.mit.edu			
Project Information			
Center project name: L5154			
Center clone name: 58_M12			
Summary Statistics			
Sequencing vector: M13; M7815; 100% of reads			
Chemistry: Dye-terminator Big Dye; 100% of reads			
Assembly program: Phrap; version 0.960731			
Consensus quality: 134743 bases at least Q40			
Consensus quality: 139227 bases at least Q30			
Consensus quality: 140814 bases at least Q20			
Insert size: 144000; agarose-fp			
Insert size: 142179; sum-of-ctnigs			
Quality coverage: 4.6 in Q20 bases; agarose-fp			
Quality coverage: 4.7 in Q20 bases; sum-of-ctnigs			
NOTE: This is a 'working draft' sequence. It currently			
consists of 10 contigs. The true order of the pieces			
is not known and their order in this sequence record is			
arbitrary. Gaps between the contigs are represented as			
runs of N, but the exact sizes of the gaps are unknown.			
This record will be updated with the finished sequence			
as soon as it is available and the accession number will			
be preserved.			
1 38820: contig of 38820 bp in length			
38821 38920: gap of 100 bp			
38921 40411: contig of 1491 bp in length			
40412 40511: gap of 100 bp			
40512 43379: contig of 2768 bp in length			
43380 43379: gap of 100 bp			
43380 46905: contig of 3526 bp in length			
46906 47005: gap of 100 bp			
47006 51830: contig of 4825 bp in length			
51831 51930: gap of 100 bp			
51931 62619: contig of 10689 bp in length			
62620 62719: gap of 100 bp			
62720 75408: contig of 12689 bp in length			

FEATURES	* 75409 75508: gap of 100 bp
	* 75509 92516: contig of 17008 bp in length
	* 92517 92616: gap of 100 bp
	* 92617 106409: contig of 13793 bp in length
	* 106410 106509: gap of 100 bp
	* 106510 143079: contig of 36570 bp in length.
SOURCE	1. 143079
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="RP11-58M12"
	/clone_lib="RPC1-11 Human Male BAC"
misc_feature	1: 38820
	/note="assembly_fragment"
	clone_end:SP6
	vector_side:left
misc_feature	38921..40411
	/note="assembly_fragment"
misc_feature	40512..43279
	/note="assembly_fragment"
misc_feature	43380..46905
	/note="assembly_fragment"
misc_feature	47006..51830
	/note="assembly_fragment"
misc_feature	51931..62619
	/note="assembly_fragment"
misc_feature	62720..75408
	/note="assembly_fragment"
misc_feature	75509..92516
	/note="assembly_fragment"
misc_feature	92617..106409
	/note="assembly_fragment"
misc_feature	106510..143079
	/note="assembly_fragment"
	clone_end:R7
	vector_side:right"
BASE COUNT	43971 a 26246 c 26678 g 45278 t 906 others
ORIGIN	
alignment_scores:	
Quality:	26.00 Length: 26
Ratio:	1.000 Gaps: 0
Percent Similarity:	100.000 Percent Identity: 100.000
alignment_block:	
US-08-973-363-8 x AC021449 ..	
Align seg 1/1 to: AC021449 from: 1 to: 143079	
1 lleleuproaspaspProaspLysLysProGlnAlaLysGlnleuGlnrh 17	
116147 ATTCTTCAGAGATGATCCGATATAAACCACCAACGAAACAGTTGCAGAC 116196	
17 rargAlaaspyrYrleuIlleLysleuLeu 26	
116197 CCGTGCAGACTACTCATCAATTACTT 116224	
seq_name: gb_htg:AC008531	
seq_documentation_block:	
LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001	
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE.	
7 ordered pieces.	
ACCESSION AC008531	
VERSION AC008531.3 GI:12830078	
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.	
SOURCE human.	
ORGANISM Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE 1 (bases 1 to 145659)	
AUTHORS DOE Joint Genome Institute.	



```
* 20435 20534: gap of unknown length
* 20535 23515: contig of 2981 bp in length
* 23516 23615: gap of unknown length
* 23616 27563: contig of 3948 bp in length
* 27564 27663: gap of unknown length
* 27664 30987: contig of 3324 bp in length
* 30988 31087: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36165: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5962 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57459: gap of unknown length
* 57460 67881: contig of 10422 bp in length
* 67882 67981: gap of unknown length
* 67982 74132: contig of 6151 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7616 bp in length
* 87512 87611: gap of unknown length
* 87612 92791: contig of 5180 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 9903 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 27465 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
FEATURES
    source          1..193446
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="Rp11-36012"
                    /clone_lib="RPCT human BAC library 11"
BASE COUNT      57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN
alignment_scores:
    Quality:      26.00      Length:      26
    Ratio:        1.000      Gaps:        0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-08-973-363-8 x AC091946 ..
Align seg 1/1 to: AC091946 from: 1 to: 193446
1 11leuProAspAspProAspPlysPlysProGlnAlaLysGlnLeuGlnth 17
|||||
87658 ATTCTTCCAGATGATCCCGATTAACCAACCAACCAACGATTGCAGAC 87707
17 rArgAlaAspTyrLeuIleLysLeuLeu 26
|||||
87708 CCGTGCAGACTACCTCATCAATTAATTACTT 87735
seq_name: gb_pr:AC026778
```

```
seq_documentation_block:
LOCUS      AC026778      195433 bp      DNA      linear      PRI 01-JUN-2001
DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.
ACCESSION  AC026778
VERSION    AC026778.4      GI:14277282
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 195433)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            TITLE      Direct Submission
            JOURNAL     Unpublished
            AUTHORS     DOE Joint Genome Institute.
            TITLE      DOE Joint Genome Institute.
            JOURNAL     Direct Submission
            AUTHORS     Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
            TITLE      Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            JOURNAL     3 (bases 1 to 195433)
            AUTHORS     DOE Joint Genome Institute and Stanford Human Genome Center.
            TITLE      Direct Submission
            JOURNAL     Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            AUTHORS     Drive, Walnut Creek, CA 94598, USA
            TITLE      On Jun 1, 2001 this sequence version replaced gi:13677045.
            JOURNAL     Draft Sequence Produced by DOE Joint Genome Institute
            COMMENT      www.jgi.doe.gov
            COMMENT      www-shgc.stanford.edu
            COMMENT      Quality: Phrap Quality >=40 99.9% of Sequence;
            COMMENT      Estimated Total Number of Errors is 0.2.
            COMMENT      SFS Content:
            COMMENT      WI-13675 G23101
            COMMENT      SHGC-58345 G38487
            COMMENT      SHGC-103595 G57841.
FEATURES
    source          1..195433
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="5"
                    /clone="CTC-428111"
BASE COUNT      62762 a 37302 c 37040 g 58329 t
ORIGIN
alignment_scores:
    Quality:      26.00      Length:      26
    Ratio:        1.000      Gaps:        0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-08-973-363-8 x AC026778/rev ..
Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433
1 11leuProAspAspProAspPlysPlysProGlnAlaLysGlnLeuGlnth 17
|||||
29732 ATTCTTCCAGATGATCCCGATTAACCAACCAACCAACGATTGCAGAC 29683
17 rArgAlaAspTyrLeuIleLysLeuLeu 26
|||||
29682 CCGTGCAGACTACCTCATCAATTAATTACTT 29655
seq_name: gb_pr:AC022121
seq_documentation_block:
LOCUS      AC022121      219258 bp      DNA      linear      PRI 30-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
ACCESSION  AC022121
VERSION    AC022121.6      GI:15375145
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
```

REFERENCE	Mamaliajoca; Metazoa; Chordata; Ctenulata; Vertebrata; Euteleostomi; Eumalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	DOE Joint genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 219258)
AUTHORS	DOE Joint genome Institute.
TITLE	Direct Submission
JOURNAL	Submitted (26-JAN-2000), Production Sequencing Facility, DOE Joint genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	3 (bases 1 to 219258)
AUTHORS	DOE Joint genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Submitted (11-AUG-2001), DOE Joint genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE	4 (bases 1 to 219258)
AUTHORS	DOE Joint genome Institute and Stanford Human Genome Center.
TITLE	Direct Submission
JOURNAL	Submitted (30-AUG-2001), DOE Joint genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT	On Aug 30, 2001 this sequence version replaced g1:15148108.

**THIS PAGE BLANK (USPTO)**





1188 TCGTGCAGATTACCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 1237  
34 luA1aGlnArGleuAlaGlyAla 41  
|||||  
1238 AAGCAGACGACCTGCTGTGTGCA 1260

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751

seq\_documentation\_block:  
ID AA142751 standard; cDNA; 6608 BP.

AA142751;

12-MAR-1997 (first entry)

Chicken CHD-1A gene.

Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;  
CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

Gallus sp.

Key Location/Qualifiers  
FH 228..5390  
FT CDS /\*tag= a

WO9639505-A1.

12-DEC-1996.

05-JUN-1996; 96WO-GB01341.

06-JUN-1995; 95GB-0011439.

(ISIS-) ISIS INNOVATION LTD.

Griffiths R, Tiwari B;

WPI; 1997-043127/04.

Avian chromodomain-helicase-DNA binding genes determine sex in  
birds - used for sex determ. and to control sex of progeny

Claim 1; Fig 5; 76pp; English.

The chicken CHD-W gene (AA142754) acting alone or in conjunction with  
the closely related CHD-1A gene (AA142751) is suggested to initiate  
female development in birds. The sequence of CDH-1A was deduced  
from 3 clones isolated from a stage 10-12 chicken embryo cDNA  
library using a great tit CHD-W sequence (see also AA142755) as probe.  
The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1  
gene (see also AA142756-57). It is located on an autosome or Z  
chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-  
specific signal on hybridisation to genomic DNA of a non-rare  
bird and can be used for sex determination of a bird. CHD-1A nucleic  
acids can also be used to control the sex of the progeny of a bird.  
Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment\_scores:  
Quality: 41.00 Length: 41  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x AA142751 ..

Align seg 1/1 to: AA142751 from: 1 to: 6608

1 lIleuPProAspPProAspLysLysProGlnAlaLysGlnLeuGlnH 17  
|||||  
4080 ATTTTACCTGATGATCCAGACAGAAACCCAGGCAAGACGACTACAGAC 4129

17 rArgAlaAspTyrLeuIlleLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
4130 CCGTGACGACTACCTCATTAATTACTGAATTAAGACCTTGCAAGAAAG 4179  
34 luA1aGlnArGleuAlaGlyAla 41  
|||||  
4180 AAGCAGAAAGCCTGCTGTGTGCA 4202

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142757

seq\_documentation\_block:  
ID AA142757 standard; DNA; 153 BP.

AA142757;

12-MAR-1997 (first entry)

Chick CHD-1A gene fragment.

Bird; sex determination; chromodomain-Helicase-DNA binding 1;  
CHD-1A; CHD-W; W chromosome; ss.

Gallus sp.

Key Location/Qualifiers  
FH misc\_difference 52..81  
FT /\*tag= a  
FT /\*note= "bases 52-81 are a repeat of bases 22-51  
and are ignored in the translated amino  
acid sequence given in Fig 3"

WO9639505-A1.

12-DEC-1996.

05-JUN-1996; 96WO-GB01341.

06-JUN-1995; 95GB-0011439.

(ISIS-) ISIS INNOVATION LTD.

Griffiths R, Tiwari B;

WPI; 1997-043127/04.

P-PSDB; AAW08147.

Avian chromodomain-helicase-DNA binding genes determine sex in  
birds - used for sex determ. and to control sex of progeny

Claim 8; Fig 3; 76pp; English.

Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology  
to portions of the chicken CHD-1A (A = Avian) gene (AA142757).  
Chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)  
and the great tit CHD-W gene (see also AA142759). Translated amino  
acid sequences of this region are provided in AAW08146-49. The  
CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes  
determine sex in birds and can be used to identify the sex of an  
embryo, foetus etc. and to manipulate the sex of progeny.  
Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment\_scores:  
Quality: 34.00 Length: 34  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x AA142757 ..

Align seg 1/1 to: AA142757 from: 1 to: 153

```

      8  LysLysProGlnAlaIysGlnLeuGlnThrArgAlaAspTyrLeuIleIle 24
      52  AAGAAACCCAGCAGCAAGACACTACAGACCCGTCAGACACTACTCATTTAA 101
      24  stLeuLysAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGln 41
      102  ATTACTGAAATAAAGACCTTGCAAGAAAGAACACAAAGCGCTTGCTGCGTG 151
      41  la 41
      152  CA 153

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758

seq_documentation_block:
ID  AA142758 standard; DNA: 153 BP.
AC  AA142758;
XX
XX  12-MAR-1997 (first entry)
XX
DE  Chick CHD-W gene fragment.
XX
XX  Bird; sex determination: chromodomain-Helicase-DNA binding 1;
XX  CHD-1a; CHD-W; W chromosome; ss.
XX
OS  Gallus sp.
XX
FH  Key Location/Qualifiers
FT  misc_difference 52..81
FT  /tag= a
FT  /note= "bases 52-81 are a repeat of bases 22-51
FT  and are ignored in the translated amino
FT  acid sequence given in Fig 3"
XX
XX  WO9639505-A1.
XX
XX  12-DEC-1996.
XX
XX  05-JUN-1996; 96WO-GB01341.
XX
XX  06-JUN-1995; 95GB-0011439.
XX
XX  (ISIS-) ISIS INNOVATION LTD.
XX
XX  Griffiths R, Tiwari B;
XX
XX  WPI: 1997-043127/04.
XX  P-PSDB: AAM08148.
XX
XX  Avian chromodomain-helicase-DNA binding genes determine sex in
XX  birds - used for sex determ. and to control sex of progeny
XX
XX  Claim 8; Fig 3; 76pp; English.
XX
XX  Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX  to portions of the chicken CHD-1A (A - Avian) gene (AA142757),
XX  chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX  and the great tit CHD-W gene (AA142759). Translated amino acid
XX  sequences of this region are provided in AAM08146-49. The CHD-1A
XX  (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX  sex in birds and can be used to identify the sex of an embryo.
XX  foetus etc. and to manipulate the sex of progeny.
XX
XX  Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

      8  LysLysProGlnAlaIysGlnLeuGlnThrArgAlaAspTyrLeuIleIle 24
      52  AAGAAACCCAGCAGCTTAAGCACTTACAGACCCGTCAGACACTACTCATTTAA 101
      24  stLeuLysAsnLysAspLeuAlaArgLysGlnAlaGlnArgLeuAlaGln 41
      102  ATTACTGAAATAAAGACCTTGCAAGAAAGAACACAGACACTTGCTGCGTG 151
      41  la 41
      152  CA 153

seq_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142759

seq_documentation_block:
ID  AA142759 standard; DNA: 153 BP.
AC  AA142759;
XX
XX  12-MAR-1997 (first entry)
XX
XX  Great tit CHD-W gene fragment.
XX
XX  Bird; sex determination: chromodomain-Helicase-DNA binding 1;
XX  CHD-1a; CHD-W; W chromosome; ss.
XX
XX  Parus major.
XX
FH  Key Location/Qualifiers
FT  misc_difference 52..81
FT  /tag= a
FT  /note= "bases 52-81 are a repeat of bases 22-51
FT  and are ignored in the translated amino
FT  acid sequence given in Fig 3"
XX
XX  WO9639505-A1.
XX
XX  12-DEC-1996.
XX
XX  05-JUN-1996; 96WO-GB01341.
XX
XX  06-JUN-1995; 95GB-0011439.
XX
XX  (ISIS-) ISIS INNOVATION LTD.
XX
XX  Griffiths R, Tiwari B;
XX
XX  WPI: 1997-043127/04.
XX  P-PSDB: AAM08149.
XX
XX  Avian chromodomain-helicase-DNA binding genes determine sex in
XX  birds - used for sex determ. and to control sex of progeny
XX
XX  Claim 8; Fig 3; 76pp; English.
XX
XX  Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX  to portions of the chicken CHD-1A (A - Avian) gene (AA142757),
XX  chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX  and the great tit CHD-W gene (AA142759). Translated amino acid
XX  sequences of this region are provided in AAM08146-49. The CHD-1A
XX  (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX  sex in birds and can be used to identify the sex of an embryo.
XX  foetus etc. and to manipulate the sex of progeny.
XX
XX  Sequence 153 BP; 60 A; 33 C; 31 G; 29 T; 0 other;

alignment_scores:

```



```

CC progeny.
XX
S0 Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

alignment_scores:
    Quality: 19.00      Length: 19
    Ratio: 1.000      Caps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000

alignment block:
US-08-973-363-8 x AAT42756 ..

Align seq 1/1 to: AAT42756 from: 1 to: 153

      8  LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleIleTyr  24
      52  AAAAAAAAAACCAAGCAAAACAGTTACAGACCGCGACGACTCATCA  101
      24  sleuteun 26
          |||||
      102  ACTACTT 108

seq_name: /SIDS1/5/cgdata/hold-geneseq/genesegn-emb1/NA2001B.DAT:AB106443

seq_documentation_block:
ID  AB106443 standard; cDNA; 6240 BP.
XX
AC  AB106443;
XX
DT  26-MAR-2002 (first entry)
DE  Drosophila melanogaster expressed polynucleotide SPQ ID NO 13811.
XX
KW  Drosophila; developmental biology; cell signalling; insecticide;
XX  pharmaceutical; gene; ss.
OS  Drosophila melanogaster.
XX
FN  WO200171042-A2.
XX
PD  27-SEP-2001.
XX
FE  23-MAR-2001; 2001WO-US09231.
XX
PR  23-MAR-2000; 2000US-191637P.
XX
PR  11-JUL-2000; 2000US-0614150.
XX
PA  (PEKE ) PE CORP NY.
XX
PI  Venter JC, Adams M, Li PWD, Myers EW;
DR  WPI: 2001-656860/75.
DR  P-PSDB: ABB62340.
XX
PT  New isolated nucleic acid detection reagent for detecting 1000 or more
PT  genes from Drosophila and for elucidating cell signalling and cell-cell
PT  interactions -
XX
XX  Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.
XX
XX  The invention relates to an isolated nucleic acid detection reagent
XX  capable of detecting 1000 or more genes from Drosophila. The invention is
XX  useful in developmental biology and in elucidating cell signalling and
XX  cell-cell interactions in higher eukaryotes for the development of
XX  insecticides, therapeutics and pharmaceutical drugs. The invention
XX  discloses genomic DNA sequences (AB1616176-AB130511), expressed DNA
XX  sequences (AB101840-AB1616175) and the encoded proteins
XX  (AB57737-AB572072).
XX  The sequence data for this patent did not form part of the printed
XX  specification, but was obtained in electronic format directly from WIPO
XX  at ftp.wipo.int/pub/published_pct_sequences.
XX

```

```

SQ Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

alignment_scores:
    Quality: 11.00      Length: 11
    Ratio: 1.000        Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-08-973-363-8 x ABL06443 ..

Align seq 1/1 to: ABL06443 from: 1 to: 6240

      9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19
      |||||
4205 AAGCCCCAGGCGAAGCAGCTGCAGACGCGTCC 4237

seq_name: /SIDS1/9cgcdata/hold-geneseq/geneseg-emb1/MA2001B.DAT:ABL06442

seq_id: ABL06442 standard; cDNA; 9933 BP.
XX
XX ABL06442;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 13608.
XX
XX Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
XX
XX P-PSDB: ABB62339.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1: SEQ ID NO 13608; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX CC sequences (AB87737-AB872072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 9933 BP; 2711 A; 2458 C; 2538 G; 2216 T; 0 other;

alignment_scores:
    Quality: 11.00      Length: 11
    Ratio: 1.000        Gaps: 0

```

Percent similarity: 100.000 Percent identity: 100.000

alignment block:  
US-08-973-363-8 x ABL06442 ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933

9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19  
|||||  
6745 AGCCCCAGCCAGCAGCTGCAGAGCGCTGCC 6777

seq\_name: /SIDS1/gcgcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK88882

seq\_documentation\_block:

ID AAK88882 standard; cDNA: 421 BP.

XX AAK88882;

XX 05-NOV-2001 (first entry)

DE Human digestive system antigen coding sequence SEQ ID NO: 1198.

KW Human: digestive system antigen; gene therapy; cancer; appendicitis;

KM ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;

KM digestive system disorder; Meckel's diverticulum; ss.

OS Homo sapiens.

PN WO200155314-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01324.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0227009.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235835.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 13-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 01-NOV-2000; 2000US-0241826.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251858.  
PR 08-DEC-2000; 2000US-0251859.  
PR 08-DEC-2000; 2000US-0251969.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM:  
XX  
DR WPI; 2001-502630/55.  
XX  
DR P-PSDB; AAM93109.  
XX  
PT Polynucleotides encoding digestive system antigens, useful for  
PT diagnosis, treating, preventing and/or prognostizing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
PS Claim 1; SEQ ID NO 1198; 986bp; English.  
XX  
CC The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a cDNA encoding a digestive  
CC system antigen of the invention.  
XX  
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;  
XX  
Alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
XX  
alignment\_block:  
US-08-973-363-8 x AAK88882 ..  
XX  
Align seg 1/1 to: AAK88882 from: 1 to: 421  
XX  
13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22  
146 AAGCAGCTACAGACCCGACGATTTACTTG 175  
XX  
seq\_name: /sids1/gcgdata/hold-geneseq/geneseg-emb1/NA2001A.DAT:AA157603  
seq\_documentation\_block:  
ID AA157603 standard; cDNA; 421 BP.  
XX  
AC AA157603;  
XX  
XX  
DT 19-OCT-2001 (first entry)  
XX  
XX Human colorectal cancer antigen cDNA SEQ ID NO: 67.

XX  
KW Human: colorectal cancer; colorectal cancer antigen; gene therapy; ss.  
XX Homo sapiens.  
OS  
XX WO20015350-A1.  
XX  
PD 02-AUG-2001.  
XX  
XX  
PE 17-JAN-2001; 2001WO-US01350.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0186874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-MAR-2000; 2000US-0198123.  
PR 19-MAR-2000; 2000US-0205315.  
PR 07-MAR-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225457.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226773.  
PR 22-AUG-2000; 2000US-0226881.  
PR 22-AUG-2000; 2000US-0226888.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.

PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236368
PR	29-SEP-2000	2000US-0236370
PR	29-SEP-2000	2000US-0236570
PR	13-OCT-2000	2000US-0239937
PR	20-OCT-2000	2000US-0240960
PR	20-OCT-2000	2000US-0241221
PR	20-OCT-2000	2000US-0241785
PR	20-OCT-2000	2000US-0241786
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	01-NOV-2000	2000US-02441826
PR	01-NOV-2000	2000US-0244617
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246603
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209
PR	17-NOV-2000	2000US-0249210
PR	17-NOV-2000	2000US-0249211
PR	17-NOV-2000	2000US-0249212
PR	17-NOV-2000	2000US-0249213
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249267
PR	17-NOV-2000	2000US-0249269
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249300
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250391
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251988
PR	05-DEC-2000	2000US-0256197
PR	06-DEC-2000	2000US-0251479
PR	08-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251868
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251989
PR	11-DEC-2000	2000US-0254097

```

XX 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
DR WPI; 2001-457727/49.
XX P-PSDB; AAM38625.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
PT disorders related to the colon and rectum including colorectal cancers
PT and also for testing and detection e.g. diagnosis -
XX
XX Claim 1; SEQ ID NO: 67; 522pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
CC number of colorectal cancer antigens. These are shown in
CC AA157547-AA157619 and AAM38569-AAM38641. These can be used in the
CC diagnosis, prevention and treatment of cancer of the colon and/or rectum
CC The present sequence is a colorectal cancer antigen coding sequence of
CC the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;
XX
XX alignment_scores:
XX      Quality: 10.00      Length: 10
XX      Ratio: 1.000      Gaps: 0
XX      Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX alignment_block:
XX US-08-973-363-8 x AA157603 ..
XX
XX Align seg 1/1 to: AA157603 from: 1 to: 421
XX
XX 13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22
XX |||||||
XX 146 AAGCAGCTACAGACCCGAGCGATTACTTG 175
XX
XX seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV33521
XX
XX seq_documentation_block:
XX ID: AAV33521 standard; cDNA; 252 BP.
XX
XX AAV33521:
XX AC
XX AT
XX 29-DEC-1998 (first entry)
XX
XX DE Clone 23789 cDNA fragment encoding a human protein kinase homolog.
XX
XX Protein kinase; cell signalling; inflammation; carcinoma; diabetes;
XX human x-linked agammaglobulinemia; nonspherocytic haemolytic anaemia;
XX artherosclerosis; glioma; restenosis; cholera-based septic shock;
XX CHKFRNK chicken tyr kinase; ss.
XX
XX Homo sapiens.
XX OS
XX US5817479-A.
XX PN
XX 06-OCT-1998.
XX PD
XX 07-AUG-1996; 96US-0700575.
XX PE
XX 07-AUG-1996; 96US-0700575.
XX PR
XX (INCY-) INCYTE PHARM INC.
XX PA
XX Au-Young J, Bandman O, Hawkins PR, Wilde CG;
XX WPI; 1998-556387/47.
XX DR

```



```

XX PT Human kinase polynucleotide(s) and recombinant products - useful
XX PT for identification of modulators of the enzyme, and treatment of
XX PT diseases associated with abnormal kinase expression
XX PS
XX PS Claim 1; Columns 29-30; 30pp; English.
XX
XX The invention provides polynucleotides which encode novel protein
XX CC kinase homologs expressed in various human cells and tissues. The
XX CC present sequence represents the clone 23789 cDNA fragment derived
XX CC from a human inflamed adenoid cDNA library. The cDNA encodes a
XX CC protein kinase which shows homology to the GHRFRK chicken Tyr kinase.
XX CC Vectors and host cells can be used for recombinant production of the
XX CC protein kinase homolog. The recombinant proteins may be used to raise
XX CC antibodies for use as anti-kinase therapeutics. Oligonucleotides based
XX CC on the polynucleotide sequences, i.e. probes and antisense constructs,
XX CC on the peptides and antibodies are claimed to be useful as tools for
XX CC studying signalling cascades in cells and proteins, and for identifying
XX CC inhibitors (drugs) to treat diseases and inflammatory conditions
XX CC associated with abnormal kinase expression. Diseases that are claimed
XX CC to be treatable include human x-linked agammaglobulinemia,
XX CC nonphagocytic haemolytic anaemia, atherosclerosis, carcinomas,
XX CC diabetes, gliomas, restenosis, cholera-toxin-induced shock, etc.
XX CC
XX Sequence 252 BP; 63 A; 70 C; 72 G; 47 T; 0 other:
XX
XX alignment_scores:
XX      Quality:      8.00      Length:      8
XX      Ratio:      1.000      Gaps:      0
XX Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX alignment_block:
XX US-08-973-363-8 x AAV33521 ..
XX
XX Align seg 1/1 to: AAV33521 from: 1 to: 252
XX
XX 24 LysleuleuAsnLysAspIleuAla 31
XX ||||||||||||||||||||||||
XX 100 AAACTCTCTACACAGACGCTGGCA 123
XX
XX seq_name: /SIDS1/9cgcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA61903
XX
XX seq_documentation_block:
XX ID ABA61903 standard: DNA; 544 BP.
XX
XX ABA61903:
XX
XX 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #10208.
XX
XX Human: foetal liver; gene expression; single exon nucleic acid probe; ss
XX
XX Homo sapiens.
XX
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001MO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207446.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0637369.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DE;
XX

```

```

XX WPI; 2001-483447/52.
DR Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX Claim 1; SEQ ID NO 10208; 639bp + sequence listing; English.
PS
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;

alignment_scores:
    Quality:      8.00          Length:      8
    Ratio:        1.000         Gaps:       0
    Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-08-973-363-8 x ABA61903 ..

Align seg 1/1 to: ABA61903 from: 1 to: 544

21 Tytleuilelyslentleuanslys 28
|||||
243 TACCTGATTAAGTTATTAAATAAG 266

seq_name: /SIBLS/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT.AAK10214
seq_documentation_block:
ID AAK10214 standard; DNA; 544 BP.
XX
AC AAK10214;
XX
DT 05-MOV-2001 (first entry)
DE Human brain expressed single exon probe SEQ ID NO: 10205.
KW Human: brain expressed exon; gene expression analysis; probe;
KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
epilepsy; cancer; ss.
XX
OS Homo sapiens.
PN MC200157275-A2.
DD
XX 09-AUG-2001.
XX
PF 30-JAN-2001; 2001MO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236559.
PR 04-OCT-2000; 2000GB-0024263.
XX
(PMOL-E) MOLECULAR DYNAMICS INC.
PA Penn Sq, Hanzel DK, Chen W, Rank DR;
XX
WPI: 2001-483446/52.
PT Single exon nucleic acid probes for analyzing gene expression in human
```

PT brains -  
XX  
PS Example 4; SEQ ID NO: 10205; 650bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system,  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-8 x AAK10214 ..  
  
Align seg 1/1 to: AAK10214 from: 1 to: 544  
  
21 TyrleuilelysleuLeuAnlys 28  
|||||  
243 TACCTGATTAAGTTATTATAATAAG 266  
  
seq\_name: /SISL1/gcgcdata/hold-geneseq/genesegn-emb1/NA2001A.DAT:AAK36112  
  
seq\_documentation\_block:  
ID AAK36112 standard; DNA; 544 BP.  
XX  
AC AAK36112;  
XX  
DF 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 10669.  
XX  
KW Human: bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
WPI; 2001-488900/53.  
XX  
DR Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human bone marrow -  
XX  
PS Example 4; SEQ ID NO: 10669; 658bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow


CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention.  
XX  
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-8 x AAK36112 ..  
  
Align seg 1/1 to: AAK36112 from: 1 to: 544  
  
21 TyrleuilelysleuLeuAnlys 28  
|||||  
243 TACCTGATTAAGTTATTATAATAAG 266

Mon Aug 5 11:52:00 2002

us-08-973-363-8.oli2n.rng

---

Page 11



**THIS PAGE BLANK (USP 10)**



DEFINITION AL644594 XGC-egg Silurana tropicalis cDNA clone L1E1D12 5', mRNA sequence.  
ACCESSION AL644594  
VERSION AL644594.1 GI:16796719  
KEYWORDS EST.  
SOURCE western clawed frog.  
ORGANISM Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Ampibbia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae; Silurana.  
REFERENCE 1 (bases 1 to 645)  
AUTHORS Huckler, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.  
TITLE Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
JOURNAL Unpublished (2001)  
COMMENT Contact: Huckler E  
Sanger Centre  
Hinnton, Cambridgeshire, CB10 1SA, UK  
Email: tropesanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: L1E1D12.pic  
Sequencing primer: PIC  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
FEATURES  
Location/Qualifiers  
1..645  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="L1E1D12"  
/clone\_lib="XGC-egg"  
/dev\_stage="egg"  
/lab\_host="Escherichia coli XL1-blue"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from egg. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end"  
BASE COUNT 222 a 125 c 156 g 141 t 1 others  
ORIGIN  
alignment\_scores:  
Quality: 30.00 Length: 30  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-8 x AL644594 ..  
Align seg 1/1 to: AL644594 from: 1 to: 645  
1 TleleuproaspaspProaspLysProGlnAlaLysGlnleuGlnrh 17  
|||||  
456 ATTTCACAGATGATCCGATTAAGAACCCCAAGCAGCAGTACAGAC 505  
17 TARGAlaAspTyrleuIleLysleuAsnLysAspleu 30  
|||||  
506 CAGAGCTGACTACCTCATTAACCTTCATTAAGATCTG 545  
seq\_name: gb\_est1:AL601246  
seq\_documentation\_block:  
LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001  
DEFINITION DKF3P313J1040\_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
DKF3P313J1040 5', mRNA sequence.  
ACCESSION AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 430)  
AUTHORS Bloeker, H., Boecher, M., Brandt, P., Mewes, W., Weil, B. and Wiemann  
S.

TITLE EST (Bloeker, H., Boecher, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloeker H  
MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
Searched by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.  
No SI sequence available.  
This clone (DKF3P313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
FEATURES  
Location/Qualifiers  
1..430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKF3P313J1040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTriplex2; Site\_1: SfiIA; Site\_2: SfiIB; cDNA-collection"  
BASE COUNT 163 a 81 c 86 g 100 t  
ORIGIN  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-8 x AL601246 ..  
Align seg 1/1 to: AL601246 from: 1 to: 430  
1 TleleuproaspaspProaspLysProGlnAlaLysGlnleuGlnrh 17  
|||||  
218 ATTCTTCAGATGATCCGATTAAGAACCAAGCAAGCAGTTCAGAC 267  
17 TARGAlaAspTyrleuIleLysleuLeu 26  
|||||  
268 CCGTGACAGACTACCTCATCAATTACTT 295  
seq\_name: gb\_est1:AI890775  
seq\_documentation\_block:  
LOCUS AI890775 547 bp mRNA linear EST 07-MAR-2000  
DEFINITION wm95f11.x1 NCI CGAP U12 Homo sapiens cDNA clone IMAGE:2443725 3',  
similar to SW:CH1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1 ; , mRNA sequence.  
ACCESSION AI890775  
VERSION AI890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 547)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.



1 l1leuProaspPProaspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
489 ATTCTCCAGATGATCTGATTAACCAAGCAAGCAAGTACAGAC 538  
17 rARGAlaAspTyrLeuLeuLeuLeu 26  
|||||  
539 CCGTGCAGACTACCTCAATCAACTACTT 566

seq\_name: gb\_est1:BB461065

seq\_documentation\_block:

LOCUS BB461065 660 bp mRNA linear EST 25-OCT-2001  
DEFINITION BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion  
Mus musculus cDNA clone D130070B13 3' similar to l10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.

ACCESSION BB461065  
VERSION BB461065.2 GI:16426612

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE  
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
'M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
'D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeeda,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)

TITLE  
JOURNAL On Jul 21, 2000 this sequence version replaced gi:9356558.

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh

'M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura

'S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and

Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

'Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamano,I., Aizawa

'K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and

Hayashizaki,Y.

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp/) for

further details.

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

FEATURES  
Source Location/Qualifiers

1. 660

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D130070B13"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site-1: Saito; Site-2: BamHI, cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGAGCGCCGCAACTCGAGATTGTTTTTTTTTNN 3'], cDNA was  
prepared by using triphase thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5',  
GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 211 a 130 c 168 g 151 t

alignment\_scores:

Quality: 26.00 Length: 26

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-8 x BB461065 ..

Align seg 1/1 to: BB461065 from: 1 to: 660

1 l1leuProaspPProaspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
481 ATTCTCCAGATGATCTGATTAACCAAGCAAGCAAGTACAGAC 530

17 rARGAlaAspTyrLeuLeuLeuLeu 26  
|||||

531 CCGTGCAGACTACCTCAATCAACTACTT 558

seq\_name: gb\_est1:AU125712

seq\_documentation\_block:

LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000

DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA

sequence.

ACCESSION AU125712

VERSION AU125712.1 GI:10950428

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 866)

Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,

Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and

Isogai,T.

HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,

Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki

'Y., Sugano,S., Isogai,T.)

Unpublished (2000)

CONTACT: Takao Isogai

Genomics Laboratory

Helix Research Institute

1533-3 Yana, Kisarazu, Chiba 292-0812, Japan

Tel: 81-438-52-3951

Fax: 81-438-52-3952

Email: genomics@hri.co.jp



HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES  
source  
Location/Qualifiers

1..866  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RM4002061"  
/clone\_id="NT2RM4"  
/cell\_type="teratocarcinoma"  
/cell\_line="NT2"  
/note="Vector: pME18SF3; mRNA from uninduced NT2 neuronal precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t 2 others  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x AU125712 ..

Align seg 1/1 to: AU125712 from: 1 to: 866

1 lileuProaspaspProaspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
450 ATTCTTCGAGATGATCCGATTAACACCAAGCAAGAGTTGACGAC 499  
17 rArgAlaaspTyrLeuIleLysLeuLeu 26  
|||||  
500 CCGTGCAGACCTACCTCATCAATTAATTCTT 527

seq\_name: gb\_est2:BE895133

seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060F1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',

ACCESSION BE895133

VERSION BE895133.1 GI:10358221

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1028)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DRP  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1469753 row: h column: 16  
High quality sequence stop: 488.

FEATURES  
source  
Location/Qualifiers

1..1028  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3921087"  
/clone\_id="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1; Note:

Site-2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 387 a 205 c 238 g 198 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x BE895133 ..

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 lileuProaspaspProaspLysProGlnAlaLysGlnLeuGlnth 17  
|||||  
131 ATTCTTCGAGATGATCCGATTAACACCAAGCAAGAGTTGACGAC 180  
17 rArgAlaaspTyrLeuIleLysLeuLeu 26  
|||||  
181 CCGTGCAGACCTACCTCATCAATTAATTCTT 208

seq\_name: gb\_est1:AW96787

seq\_documentation\_block:

LOCUS AW96787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW96787

VERSION AW96787.1 GI:8257021

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 337)

AUTHORS Dias Neto E., Garcia Correa R., Verjovski-Almeida S., Briones M.R.,

Nagai M.A., da Silva W. Jr., Zago M.A., Bordin S., Costa F.F.,

Goldman G.H., Carvalho A.F., Mitsuoka A., Bala G.S., Simpson D.H.,

Brunstein A., deoliveira P.S., Bucher P., Jongeneel C.V., O'Hare

M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and

Simpson A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

CONTACT: Simpson A.J.C.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=42-QV3-BN0047-230>)

Seq primer: puc 18 forward

High quality sequence start: 2

High quality sequence stop: 337.

FEATURES  
source  
Location/Qualifiers

1..337  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="BN0047"  
/db\_xref="taxon:9606"  
/note="Organ: breast; normal; Vector: puc18; Site: 1; Smal;  
Site: 2; Smal; A mini-library was made by cloning products  
derived from ORESSES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

ORIGIN

alignment\_scores:

Quality: 23.00 Length: 23  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-8 x AM996787/rev ..

Align seg 1/1 to reverse of: AM996787 from: 1 to: 337

1 ttleupProaspProraaplylsPrognAlaLysGlnleuglnh 17  
|||||  
105 ATTCTTCAGATGATCCGATATAAAACCAAGCAAAACAGTTGCAGAC 56

17 rArgAlaasPTyrLeuile 23  
|||||  
55 CCGTGCAGACTATCTCATC 37

seq\_name: gb\_est1:BB830730

seq\_documentation\_block:

LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527  
Jy9-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA  
sequence.

ACCESSION BB830730  
VERSION BB830730.1 GI:17008973

KEYWORDS

EST.

ORGANISM

house mouse.  
Mus musculus

REFERENCE

1 (bases 1 to 438)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Ishii Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL

CONTACT: Yoshinide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh

M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watanishi,M., Tameda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura

S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and

Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

e mouse tissues.

FEATURES Location/Qualifiers

source

1. 438  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930013K04"  
/clone\_11b="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jy9-MC(B)"

BASE COUNT

138 a 96 c 108 g 96 t

alignment\_scores:

Quality: 22.00 Length: 22  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-8 x BB830730 ..

Align seg 1/1 to: BB830730 from: 1 to: 438

1 ttleupProaspProraaplylsPrognAlaLysGlnleuglnh 17  
|||||  
300 ATTCTTCAGATGATCCTGATATAAAACCAAGCAAAACAGTTACAGAC 349

17 rArgAlaasPTyrLeu 22  
|||||  
350 CCGTGCAGACTATCTCATC 365

seq\_name: gb\_est1:BB834922

seq\_documentation\_block:

LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001  
DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527  
Jy9-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA  
sequence.

ACCESSION BB834922  
VERSION BB834922.1 GI:17013165

KEYWORDS

EST.

ORGANISM

house mouse.  
Mus musculus

REFERENCE

1 (bases 1 to 446)  
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Ishii Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watabiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

JOURNAL

CONTACT: Yoshinide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

Waghi, K., Fujiwara, S., Inoue, K., Togawa, Y., Itawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

#### FEATURES

source 1. 446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G93003j21"  
/clone\_1b="RIKEN full-length enriched, mammary gland  
RCB-0527 JyG-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 JyG-MC(B)"

BASE COUNT 139 a 99 c 108 g 100 t

ORIGIN

alignment\_scores:  
Quality: 22.00 Length: 22  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-08-973-363-8 x BB834922 ..  
Align seg 1/1 to: BB834922 from: 1 to: 446

1 tleleuproaspapraproaspaplysProGlnAlaLysGlnLeuGlnth 17.  
|||||  
309 ATTCTCCAGATGATCTGATTAACCAACCAACCAACGTTACAGAC 358  
|||||  
17 tATGAlaAspTYrLeu 22  
|||||  
359 CCGTCAGACTACTCTC 374

seq\_name: gb-est2:BF239967

seq\_documentation\_block:  
LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000

DEFINITION 601965170P1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5', mRNA sequence.

ACCESSION BF239967

VERSION BF239967.1 GI:11153890

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

TITLE NIH-MGC <http://mgi.mgi.nih.gov/>

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMU)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMU at:  
<http://image.llnl.gov>  
Plate: LLCM1033 row: k column: 18  
High quality sequence stop: 562.

#### FEATURES

source 1. 821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4133129"  
/clone\_1b="NIH\_MGC\_54"  
/tissue\_type="from chronic myelogenous leukemia"  
/lab\_host="DH10B (TI phage-resistant)"  
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech); Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggccatagcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor (where B = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 345 a 149 c 190 g 137 t

ORIGIN

alignment\_scores:  
Quality: 21.00 Length: 21  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-08-973-363-8 x BF239967 ..  
Align seg 1/1 to: BF239967 from: 1 to: 821

6 ProAspLysLysProGlnAlaLysGlnLeuGlnthArgAlaAspTYrle 22  
|||||  
3 CCGGATTAATAAAACCAACCAACCAACCAACCGTTCAGACCGTTCAGACTACT 52  
|||||  
22 uilleLysLeuLeu 26  
|||||  
53 CATCAAAATTACTT 65

seq\_name: gb-est1:AW97058

seq\_documentation\_block:  
LOCUS AW97058 686 bp mRNA linear EST 05-JUN-2000

DEFINITION OVA-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.

ACCESSION AW97058

VERSION AW97058.1 GI:8257292

KEYWORDS EST

SOURCE human

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

TITLE Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,

JOURNAL Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,

MEDLINE Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

COMMENT Brunslein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL:  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=4t2-QV3-BN0047-150400-152-C03&t3=2000-04-15&t4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 19  
 High quality sequence stop: 678.  
 Location/Qualifiers  
 1..686  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="BN0047"  
 /dev\_stage="Adult"  
 /note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 165 a 154 c 126 g 241 t  
 ORIGIN

alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-8 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

9 LysPProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuLeuLysLe 25  
 |||  
 104 AAACCAAGCAACAGTTCAGACCGTGCAGACTCACTCAACAAATT 55

25 ulen 26  
 |||  
 54 ACTT 51

seq\_name: gb\_gss:CNS04DVG

seq\_documentation\_block:

LOCUS CNS04DVG 856 bp DNA linear GSS 21-MAY-2000  
 DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone  
 103K08 of library G from Tetradon nigroviridis, genomic survey  
 sequence.

ACCESSION AL286261.1 GI:8024707  
 VERSION AL286261.1  
 KEYWORDS GSS: genome survey sequence.  
 SOURCE Tetradon nigroviridis.  
 ORGANISM Tetradon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorphi; Acanthopterygii; Percomorphi; Tetraodontiformes;  
 Tetraodontidae; Tetradon.

1 (bases 1 to 856)  
 Roest-Crollius,H., Jallou,O., Dasilva,C., Fizeses,C., Fisher,C.,  
 Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and  
 Weissenbach,J.  
 Characterization and repeat analysis of the compact genome of the  
 freshwater pufferfish Tetradon nigroviridis

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 856)  
 Roest-Crollius,H., Jallou,O., Dasilva,C., Bouneau,L., Fisher,C.,  
 Bernot,A., Fizeses,C., Wincker,P., Brothier,P., Quetier,F.,

TITLE Saurin,W. and Weissenbach,J.  
 JOURNAL Human gene number estimate provided by genome wide analysis using  
 REFERENCE Tetradon nigroviridis DNA sequence  
 AUTHORS Unpublished  
 TITLE 3 (bases 1 to 856)  
 JOURNAL Genoscope.  
 COMMENT Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases  
 This sequence is a single read and was generated as part of a large  
 scale clone-end sequencing project of the Tetradon nigroviridis  
 genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/tetradon>.

FEATURES  
 source  
 1..856  
 /organism="Tetradon nigroviridis"  
 /db\_xref="taxon:99883"  
 /clone\_lib="103K08"  
 /clone\_lib="G"  
 /note="Genoscope sequence ID : C0RG103BF04LP1-end : 77"  
 BASE COUNT 84 a 308 c 252 g 206 t 6 others  
 ORIGIN

alignment\_scores:  
 Quality: 18.00 Length: 18  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-8 x CNS04DVG/rev ..

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

2 LeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnThrAr 18  
 |||  
 674 CTCGCCGATGATCTCTGATAGAGAGCTCAGCCCAACAGTTACAGACCA 625

18 gala 19  
 |||  
 624 AGCC 621

seq\_name: gb\_est2:BF881342

seq\_documentation\_block:

LOCUS BF881342 129 bp mRNA linear EST 17-JAN-2001  
 DEFINITION PMO-ET0208-031200-001-f12 ET0208 Homo sapiens CDNA, mRNA sequence.  
 ACCESSION BF881342  
 VERSION BF881342.1 GI:12271468  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 129)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Brites,M.R.,  
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.F.

Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags  
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 20202663  
 Contact: Simpson A.J.F.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil

Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PMO&t2=PMO-ET0208-031200-001-f12&t3=2000-12-03&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 22  
High quality sequence stop: 128.

FEATURES  
Location/Qualifiers  
1..129

source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="ET0208"  
/dev\_stage="Adult"  
/note="Organ: lung\_tumor; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."  
BASE COUNT 21 a 48 c 23 g 37 t  
ORIGIN

alignment\_scores:  
Quality: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-8 x BF881342/rev ..

Align seg 1/1 to reverse of: BF881342 from: 1 to: 129

13 LysGlnLeuGlnThrAlaGlnAspTyrLeu 22  
|||||  
108 AACGACGTACAGCCGACGCGGATTTG 79

**THIS PAGE BLANK (USPTO)**



COMMENT Other publication AU 5906996 961224.  
FEATURES Location/Qualifiers  
Source 1. 1316  
/db\_xref="taxon:32644"  
BASE COUNT 493 a 205 c 308 g 304 t 6 others  
ORIGIN

alignment\_scores:  
Quality: 34.00 Length: 34  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x A58696 ..

Align seg 1/1 to: A58696 from: 1 to: 1316

1 11leuProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
1138 ATTTACTGATGATCCAGATTAAGAAACCCAGCGTACAGCTACAGAC 1187  
17 TATGALASPtyrLeuLeuLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||  
1188 TCGTGCAGATTACTCATTAATTACTGAATTAAGACCTTGCAAGAAGG 1237  
34 Lu 34  
||  
1238 AA 1239

seq\_name: gb\_pat:A58691

seq\_documentation\_block:

LOCUS A58691 6608 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 10 from Patent WO9639505.  
ACCESSION A58691  
VERSION A58691.1 GI:3714250

KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 6608)  
AUTHORS Griffiths, R. and Tiwari, B.

TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN  
BIRDS

JOURNAL Patent: WO 9639505-A 10 12-DEC-1996;

COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.

Source 1. 6608  
Location/Qualifiers  
/db\_xref="taxon:32644"

BASE COUNT 2289 a 1207 c 1459 g 1653 t  
ORIGIN

alignment\_scores:  
Quality: 34.00 Length: 34  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x A58691 ..

Align seg 1/1 to: A58691 from: 1 to: 6608

1 11leuProaspaspProaspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
4080 ATTTACTGATGATCCAGACAAAGAAACCCAGCGTACAGCTACAGAC 4129  
17 TATGALASPtyrLeuLeuLysLeuAsnLysAspLeuAlaArgLysG 34  
|||||

4130 CCGTGCAGACTACTCATTAATTACTGAATTAAGACCTTGCAAGAAAGG 4179  
34 Lu 34  
||  
4180 AA 4181

seq\_name: gb\_ov:AF004397

seq\_documentation\_block:

LOCUS AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
DEFINITION Gallus gallus chromo-helicase-DNA-binding on the Z chromosome  
protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete  
cds.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845

KEYWORDS  
SOURCE chicken.

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 6872)

AUTHORS Griffiths, R. and Korn, R.M.

TITLE A CHD1 gene is Z chromosome linked in the chicken Gallus domesticus

MEDLINE Gene 197 (1-2), 225-229 (1997)

REFERENCE 97473516

AUTHORS Griffiths, R. and Korn, R.M.

TITLE Direct Submission

Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 800, UK

FEATURES Location/Qualifiers

Source 1. 6872

gene /organism="Gallus gallus"

CDS /db\_xref="taxon:9031"

1. 6872

228. 5654

/gene="CHD-Z"

/function="role in chromatin architecture"

/note="CHD protein with hydrophilic domain"

/product="chromo-helicase-DNA-binding on the Z chromosome  
protein"

/protein\_id="AAC60282.1"

/db\_xref="GI:2501846"

/translation="MNGHSDERSVNRSGESRSDDDSAGSGSGSGSGSGSGSS  
SOSGSDSESGSGSGSSESSTSRERKQVAKPKKADGSEFMKSPSILAVORSAY  
LKROOQOKAAASDSDGSESDSSSDSADSSSRTKKKHKDEDMQSGSGSGSGS  
DSESAEDGKSSCESESDYEPKKNKYSKPPSRILKPKSGKSTGOKRQDSEED  
DDDEDYDKRGSRRQAVNYSYKAEETKTSDDLLKVGCEGDPORTEDEFEET  
SRIGKKGATGASTTYAVADGDGDPNAGPEKSKSELEIOLYLKWKMSIHHNWEET  
LKOQVKGANKLNDTKKKDDQETKRWLKNASPEDVETYNCOQELTDLHKQYVERIT  
AHSNOKSAGYDYYCKWQGLPYSECSDWGLLAKKQARIDEFNRQSKTTPPKD  
CKYLKORPERFVALRKOPSYTGGHESLELDYOLNOLNMLAHMCWGNSGTLADENGIG  
KITQITSEFNTYLFHEHOLYGPFLRLPSTLSMORLQTMAPQNNAAVYLDLSRN  
MIRTHBMHPQTKRLKFNLTLLTTEYLKDKSFLGGLMNAFVIGVDEARLKKDGLYAS  
RTLIDKSNHRLITGTPLONSLKEIMSLHIMPEKSSPEDEEBCGKGEVYAS  
LHKLELPELRLRVKQDVEKSLPAKVEQLIRMBMSLKOQYKMITLTRYKALSKSKG  
STSGFLNIMMELKRCNCNYLIRPPDNEFYKQALDHLIRSSGLLILDLILRLR  
ERGNRYLVIFESQWRMIDILAEYLKROFPFLDGSIGELKKQALDHRNAGSDPC  
FLSTIRAGIGLNLASADTVVIFSDMNPQNDLQARAHRIQOKQVNIYILVFGS  
VEDILERRAKKMKVLDHLYVIOEMDTGKTVLHTGSPSSPPEKKEELAILFGAE  
LEKPEGEQEPQEDIDELIKRAFTRENERGPLVGDILLSQFANVNSMDEDEIE  
LEPERNSRWMEETIPESQRRILEERERQLELEITMLPRMRCACQISNGBGRSR  
SNRYSGSDSDSITERRKRPKRGRPTIPRENKIGTSDAIERIKSYKFKGFLRLD  
AARDAELVYDSEKTDLRLGELVHNGCIKALDNSGQERAGRGKVGKGFPRISGV  
QVNAKLIVISHEELAPLHKSISSDEERKRYIYIPHTAAAHFDIMQKEDDSNLLVGI  
YKYGSGMELIKMDPLSTOKLIPDDDKRQAKLOTRAYLYLKLINKDLAREAO  
RLAGAGNSKRRKTRKKNKMKASKIKELIKSSSPSPSEKSEDEEDENKELISYK  
HLHAKIKTEKENEKPEPDIGITKAEERETKENRELAKEKEDDKELKED  
NKEKRENVKSESTOKKEVKEKVENMSSENKESKKLPLDTPVHTATYSPPVISE  
ESELHOKTFESVCKERMRPVKALKOLDRPEKGLSERQLEHTROCLIKIGDHTECL





[illegible][illegible]

```

1 11leuProaspaspProaspLysPProGlnAlaLysGlnLeuGlnTh 17
|||||
4025 ATTCTTCCAGATGATCCCGATTAACCAACGACAAACAGTTGCAGAC 4074
17 rArgAlaAspTyrLeuIleLysLeuLeu 26
|||||
4075 CCGTGCAGACTACCTCATCAATTAATTACTT 4102

seq_name: gb_pr:AC092372

seq_documentation_block:
LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.
ACCESSION AC092372
VERSION AC092372.3 GI:17402768
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 101220)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 101220)
TITLE Direct Submission
AUTHORS Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 101220)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 7, 2001 this sequence version replaced gi:15290448.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.hgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 146.7kb). It is clipped at the overlap with AC012624.
The number of bases overlapped is 90404.
FEATURES
source
1..101220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-58M12"
BASE COUNT 34122 a 18862 c 17827 g 30409 t
ORIGIN

alignment_scores:
Quality: 26.00 Length: 26
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-9 x AC092372/rev ..

Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220

1 11leuProaspaspProaspLysPProGlnAlaLysGlnLeuGlnTh 17
|||||
26932 ATTCTTCCAGATGATCCCGATTAACCAACGACAAACAGTTGCAGAC 26883
17 rArgAlaAspTyrLeuIleLysLeuLeu 26
|||||
26882 CCGTGCAGACTACCTCATCAATTAATTACTT 26855

seq_name: gb_pr:AC012624

```

```

seq_documentation_block:
LOCUS AC012624 134365 bp DNA linear PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2082117, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:14993679
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 134365)
TITLE Direct Submission
AUTHORS Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
source
1..134365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CTD-2082117"
BASE COUNT 40414 a 24497 c 25503 g 43951 t
ORIGIN

alignment_scores:
Quality: 26.00 Length: 26
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-9 x AC012624 ..

Align seg 1/1 to: AC012624 from: 1 to: 134365

1 11leuProaspaspProaspLysPProGlnAlaLysGlnLeuGlnTh 17
|||||
118247 ATTCTTCCAGATGATCCCGATTAACCAACGACAAACAGTTGCAGAC 118296
17 rArgAlaAspTyrLeuIleLysLeuLeu 26
|||||
118297 CCGTGCAGACTACCTCATCAATTAATTACTT 118324

seq_name: gb_htg:AC021449

seq_documentation_block:
LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
pieces.
ACCESSION AC021449
VERSION AC021449.3 GI:10047806
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 143079)

```

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens, clone RP11-58M12  
Unpublished  
2 (bases 1 to 143079)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,  
Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G., Castle, A.,  
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
DeRellano, K., Dewar, K., Domino, C., Doyle, M., Feneator, J.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardina, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karats, A., Klein, J.,  
Lander, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,  
Macdonald, P., Marquis, N., McEwan, P., McGuirk, A., McKernan, K.,  
McPheeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
Zimmer, A. and Zody, M.

TITLE  
JOURNAL

## COMMENT

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced g1:7407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

## Project Information

Center project name: L5154

Center clone name: 58\_M12

## Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 134743 bases at least Q40

Consensus quality: 139227 bases at least Q20

Consensus quality: 140814 bases at least Q20

Insert size: 144000; agarose-fp

Insert size: 142179; sum-of-contigs

Quality coverage: 4.6 in Q20 bases; agarose-fp

Quality coverage: 4.7 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 38820: contig of 38820 bp in length  
\* 38821 38820: gap of 100 bp  
\* 38921 40411: contig of 1491 bp in length  
\* 40412 40511: gap of 100 bp  
\* 40512 43279: contig of 2768 bp in length  
\* 43280 43379: gap of 100 bp  
\* 43380 46905: contig of 3526 bp in length  
\* 46906 47005: gap of 100 bp  
\* 47006 51830: contig of 4825 bp in length  
\* 51831 51930: gap of 100 bp  
\* 51931 62619: contig of 10689 bp in length  
\* 62620 62719: gap of 100 bp  
\* 62720 75408: contig of 12689 bp in length  
\* 75409 75508: gap of 100 bp  
\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length

\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.  
Location/Qualifiers  
1. 143079

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-58M12"  
/clone\_lib="RP11 Human Male BAC"  
1. 38820  
/note="assembly-fragment  
clone\_end:SP6  
vector\_side:left"

misc\_feature  
38921..40411  
/note="assembly-fragment"

misc\_feature  
40512..43279  
/note="assembly-fragment"

misc\_feature  
43380..46905  
/note="assembly-fragment"

misc\_feature  
47006..51830  
/note="assembly-fragment"

misc\_feature  
51931..62619  
/note="assembly-fragment"

misc\_feature  
62720..75408  
/note="assembly-fragment"

misc\_feature  
75509..92516  
/note="assembly-fragment"

misc\_feature  
92617..106409  
/note="assembly-fragment"

misc\_feature  
106510..143079  
/note="assembly-fragment  
clone\_end:17  
vector\_side:right"

misc\_feature  
106510..143079  
/note="assembly-fragment  
clone\_end:17  
vector\_side:right"

BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AC021449 ..

Align seg 1/1 to: AC021449 from: 1 to: 143079

1 11leuProaspaspyroaspysylsproglnalalysglnleuinh 17  
|||||

116147 ATTCTTCAGATGATCCGATTAACCAAGCAACAGATTGCAGAC 116196  
|||||

17 TATGATAASPTyrleuileuysleu 26  
|||||

116197 CCGTGCAGACTACTCTCAATTACTT 116224  
|||||

seq\_name: gb\_hrg:AC008531

seq\_documentation\_block:

LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001

DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,

7 ordered pieces.

ACCESSION AC008531

VERSION AC008531.3 GI:12830078

KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 145659)

AUTHORS DOE Joint Genome Institute.

TITLE Sequencing of Human Chromosome 5

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 145659)

AUTHORS DOE Joint Genome Institute.



*	27564	27663: gap of unknown length
*	27664	30987: contig of 3324 bp in length
*	30988	31087: gap of unknown length
*	31088	36065: contig of 4978 bp in length
*	36066	36165: gap of unknown length
*	36166	40978: contig of 4813 bp in length
*	40979	41078: gap of unknown length
*	41079	45663: contig of 4585 bp in length
*	45664	45763: gap of unknown length
*	45764	51745: contig of 5982 bp in length
*	51746	51845: gap of unknown length
*	51846	57359: contig of 5514 bp in length
*	57360	57459: gap of unknown length
*	57460	67881: contig of 10422 bp in length
*	67882	67981: gap of unknown length
*	67982	74132: contig of 6151 bp in length
*	74133	74232: gap of unknown length
*	74233	79795: contig of 5563 bp in length
*	79796	79895: gap of unknown length
*	79896	87511: contig of 7616 bp in length
*	87512	87611: gap of unknown length
*	87612	92761: contig of 5180 bp in length
*	92762	92861: gap of unknown length
*	92862	102794: contig of 9903 bp in length
*	102795	102894: gap of unknown length
*	102895	110866: contig of 7972 bp in length
*	110867	110966: gap of unknown length
*	110967	117571: contig of 6605 bp in length
*	117572	117671: gap of unknown length
*	117672	123378: contig of 6067 bp in length
*	123733	123838: gap of unknown length
*	123839	130583: contig of 6745 bp in length
*	130584	130683: gap of unknown length
*	130684	141544: contig of 10861 bp in length
*	141545	141644: gap of unknown length
*	141645	169109: contig of 27465 bp in length
*	169110	169209: gap of unknown length
*	169210	193446: contig of 24237 bp in length

```

FEATURES
source      Location/Qualifiers
1. 193446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-360J2"
/clone_1lb="RP11-360J2 human BAC library 11"
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN

```

```
alignment_scores:      26.00      Length:      25
                      quality:
                      Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
```

alignment\_block:

Align seg 1/1 to: AC091946 from: 1 to: 193446

1 IleuPProASPASPProASPylsYsProGlnAlaYsGlnIleuGlnIth 17  
87658 ATTCTTCAGATGATGCCGATPAAAAACACACAGCAAAAGTTGCAGAC 87707  
17 TArgAlaAspTyrIleuIleYsIleuLeu 26  
CGGTGCAGACATACCATCATTAATTCCTT 87735  
87708

seq\_name: gb\_pr:AC026778

seq. documentation link:	
LOCUS	195433 bp DNA PRI 01-JUN-2001
AC026778	linear
DEFINITION	Homo sapiens chromosome 5 clone CTC-42811, complete sequence.
ACCESSION	AC026778

VERSION	AC026778.4	GI:14277282
KEYWORDS	HMG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 195433)	
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.	
TITLE	Direct Submission	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 195433)	
AUTHORS	DOE Joint Genome Institute.	
TITLE	Direct Submission	
JOURNAL	Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
REFERENCE	3 (bases 1 to 195433)	
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
COMMENT	On Jun 1, 2001 this sequence version replaced gi:13677045.	

[www.jhi.doe.gov](http://www.jhi.doe.gov)  
 Fishing Completed at Stanford Human Genome Center  
[www.shgc.stanford.edu](http://www.shgc.stanford.edu)  
 Quality: Phrap Quality >=40 99.9% of Sequence;  
 Estimated Total Number of Errors is 0.2.  
 SIS Content:  
 WI-13675 G23101  
 SHGC-58345 G38487  
 SHGC-103595 G57841.

FEATURES	source	Location/Qualifiers
	1..195433	
	/Organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/Chromosome="5"	
	/Clone="CPC-428111"	
BASE COUNT	62762 a 37302 c 37040 g 58329 t	
ORIGIN		

```
alignment_scores:      26.00      Length: 26
                      quality:      Gaps: 0
                      Ratio: 1.000
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment\_block:  
US-08-973-363-9 x AC026778/rev ..

Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433

1 ILEUProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
29732 ATTCTTCAGATGATCCCGATAAAAAACCAAGCAAAACAGTTGCAGAC 29683

17 rargAlasptyrLeuIleLysLeuLeu 26

29682 CCGTGCAGACTACCTCATCAATTACTT 29655

seq_name:	gb_pr:AC022121
seq_documentation_block:	
LOCUS	AC022121 219258 bp DNA linear PRI 30-AUG-2001
DEFINITION	Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
ACCESSION	AC022121
VERSION	AC022121.6 GI:15375145

**SOURCE** human

**REFERENCE**  
**AUTHORS**  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 219258)  
 DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 219258)  
REFERENCE Direct Submission  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 219258)  
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS Direct Submission  
TITLE Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
JOURNAL On Aug 30, 2001 this sequence version replaced g1:15148108.  
COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.4.  
SPS Content:  
WT-5811 G04674  
WT-13675 G23301  
SHGC-58345 G38467  
SHGC-103595 G37841.

FEATURES  
source Location/Qualifiers  
1..219258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2007H13"

BASE COUNT 71954 a 42062 c 40933 g 64309 t  
ORIGIN

alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AC022121/rev ..

Align seg 1/1 to reverse of: AC022121 from: 1 to: 219258

1 lleleuproaspaprophasplysproglalalyglnleuglnh 17  
|||||  
89450 ATTCTTCCAGATGATCCGATATAAACCAACACAAACAGATTGCAGAC 89401  
17 TATGATAAPTATYtleuulleysleu 26  
|||||  
89400 CGGTGAGACTACTCATCAATATTCATT 89373







```

      8 LysylsProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIleLy 24
      |||
      52 AAGAAACCAAGCAAGCAAGTTCAGACCCGTCAGATTACCTCATTTAA 101
      |||
      24 sleuLeuAsnLysAspleuAlaArgLysGlnValGlnArgLeuThrGlyA 41
      |||
      102 ATTACTGTAATTAAGACCTTCGCAAGAAAAGAAAGTCAAAAGACTTACTGTTG 151
      ||
      41 la 41
      ||
      152 CA 153

seq.name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142754
seq_documentation_block:
ID   AA142754 standard; cDNA; 1316 BP.
XX
XX   AA142754;
XX
XX   12-MAR-1997 (first entry)
XX
XX   Chicken CHD-W gene (partial sequence).
XX
XX   Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
XX   CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX
XX   Gallus sp.
XX
XX   WO9639505-A1.
XX
XX   12-DEC-1996.
XX
XX   05-JUN-1996; 96WO-GB01341.
XX
XX   06-JUN-1995; 95GB-0011439.
XX
XX   (ISIS-) ISIS INNOVATION LTD.
XX
XX   Griffiths R, Tiwari B;
XX
XX   WPI; 1997-043127/04.
XX
XX   Avian chromodomain-helicase-DNA binding genes determine sex in
XX   birds - used for sex determ. and to control sex of progeny
XX
XX   Claim 1; Fig 8; 76pp; English.
XX
XX   The chicken CHD-W gene (AA142754) acting alone or in conjunction with
XX   the closely related CHD-1A gene (AA142751) is suggested to initiate
XX   female development in birds. The sequence of CDH-W was deduced
XX   from 2 clones isolated from a 10-day chick embryo library using
XX   a fragment of the CHD-1A gene as a probe. The CHD-W gene is
XX   located on the W chromosome. Probes based on CHD-W and CHD-1A give
XX   a W chromosome-specific signal on hybridisation to genomic DNA of a
XX   non-ratite bird and can be used for sex determin. of a bird. CHD-W
XX   nucleic acids can also be used to control the sex of progeny of a
XX   bird.
XX
XX   Sequence 1316 BP; 492 A; 208 C; 306 G; 304 T; 6 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-9 x AA142754 ..
Align seg 1/1 to: AA142754 from: 1 to: 1316
1 TleuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnThr
|||||
```

```

      1138 ATTTTACCTGATGATCCAGATAGAAACCCGAGCTAGACATTACAGAC 1187
      |||
      17 FAAGAlaAspTyrLeuIleLysleuAsnLysAspleuAlaArgLysG 34
      |||
      1188 TCGTGCAGATTACCTCATTTAATTACTGTAATTAAGACCTTCAGAAAGG 1237
      ||
      34 lu 34
      ||
      1238 AA 1239

seq.name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142751
seq_documentation_block:
ID   AA142751 standard; cDNA; 6608 BP.
XX
XX   AA142751;
XX
XX   12-MAR-1997 (first entry)
XX
XX   Chicken CHD-1A gene.
XX
XX   Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;
XX   CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.
XX
XX   Gallus sp.
XX
XX   Key Location/Qualifiers
XX   CDS 228..5390
XX   FT /*tag= a
XX
XX   WO9639505-A1.
XX
XX   12-DEC-1996.
XX
XX   05-JUN-1996; 96WO-GB01341.
XX
XX   06-JUN-1995; 95GB-0011439.
XX
XX   (ISIS-) ISIS INNOVATION LTD.
XX
XX   Griffiths R, Tiwari B;
XX
XX   WPI; 1997-043127/04.
XX
XX   Avian chromodomain-helicase-DNA binding genes determine sex in
XX   birds - used for sex determ. and to control sex of progeny
XX
XX   Claim 1; Fig 5; 76pp; English.
XX
XX   The chicken CHD-W gene (AA142754) acting alone or in conjunction with
XX   the closely related CHD-1A gene (AA142751) is suggested to initiate
XX   female development in birds. The sequence of CDH-W was deduced
XX   from 3 clones isolated from a stage 10-12 chicken embryo cDNA
XX   library using a great tit CHD-W sequence (see also AA142755) as probe.
XX   The CHD-1A (A = Avian) gene shows close identity to the mouse CHD-1
XX   gene (see also AA142756-57). It is located on an autosome or Z
XX   chromosome. Probes based on CHD-W and CHD-1A give a W chromosome-
XX   specific signal on hybridisation to genomic DNA of a non-ratite
XX   bird and can be used for sex determin. of a bird. CHD-1A nucleic
XX   acids can also be used to control the sex of the progeny of a bird.
XX
XX   Sequence 6608 BP; 2289 A; 1207 C; 1459 G; 1653 T; 0 other;

alignment_scores:
Quality: 34.00 Length: 34
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-9 x AA142751 ..
Align seg 1/1 to: AA142751 from: 1 to: 6608
```

```

1 lileuProaspaspProaspLySpProGlnAlaLySGlnLeuGlnH 17
|||||
4080 ATTTACTCTGATGATCCAGACAGAAACCCAGCAGCAGCTACAGAC 4129
|||||
17 rrrgAlaSpLyLeuileLyLeuAlaSpLyLeuAlaSpLyLeu 34
|||||
4130 CGTGCACAGCTACCTCATTAATTAATTAATTAAGCTTCAGAAAG 4179
|||||
34 lu 34
4180 AA 4181

seq.name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142757

seq.documentat_block:
ID AA142757 standard; DNA; 153 BP.
AC AA142757;
XX
XX 12-MAR-1997 (first entry)
DT
XX
XX Chick CHD-1A gene fragment.
DE
XX
XX Bird; sex determination; Chromodomain-Helicase-DNA binding 1;
KM CHD-1A; CHD-W; W chromosome; ss.
XX
XX Gallus sp.
OS
XX
XX Key Location/Qualifiers
FH misc_difference 52..81
FT /tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX
XX P-PSDB: AAM08147.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A - Avian) gene (AA142757)
XX chicken CHD-W (W refers to the W chromosome) gene (see also AA142758)
XX and the great tit CHD-W gene (see also AA142759). Translated amino
XX acid sequences of this region are provided in AAM08146-49. The
XX CHD-1A (see also AA142751) and CHD-W (see also AA142754-55) genes
XX determine sex in birds and can be used to identify the sex of an
XX embryo, foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 58 A; 40 C; 31 G; 24 T; 0 other;

alignment_scores:
Quality: 27.00 Length: 27
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

```

Alignment_block:
us-08-973-363-9 x AA142757
Align seg 1/1 to: AA142757 from: 1 to: 153

8 lvsLyProGlnAlaLySGlnLeuGlnHrrrAlaSpLyLeuileLy 24
|||||
52 AAGAAACCCAGCAGCAAGAGCTACAGACCCGTCAGACTACCTCATTA 101
|||||
24 slcLeuAlaSpLyLeuAlaSpLyLeuAlaSpLyLeu 34
|||||
102 ATTTACTGATTAAGCTTCAGAAAGGAA 132

seq.name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142758

seq.documentat_block:
ID AA142758 standard; DNA; 153 BP.
AC AA142758;
XX
XX 12-MAR-1997 (first entry)
DT
XX
XX Chick CHD-W gene fragment.
DE
XX
XX Bird; sex determination; Chromodomain-Helicase-DNA binding 1;
KM CHD-1A; CHD-W; W chromosome; ss.
XX
XX Gallus sp.
OS
XX
XX Key Location/Qualifiers
FH misc_difference 52..81
FT /tag= a
FT /note= "bases 52-81 are a repeat of bases 22-51
FT and are ignored in the translated amino
FT acid sequence given in Fig 3"
XX
XX MO9639505-A1.
XX
XX 12-DEC-1996.
XX
XX 05-JUN-1996; 96MO-GB01341.
XX
XX 06-JUN-1995; 95GB-0011439.
XX
XX (ISIS-) ISIS INNOVATION LTD.
XX
XX Griffiths R, Tiwari B;
XX
XX WPI: 1997-043127/04.
XX
XX P-PSDB: AAM08148.
XX
XX Avian chromodomain-helicase-DNA binding genes determine sex in
XX birds - used for sex determ. and to control sex of progeny
XX
XX Claim 8; Fig 3; 76pp; English.
XX
XX Bases 3855-3977 (AA142756) of the mouse CHD-1 gene show homology
XX to portions of the chicken CHD-1A (A - Avian) gene (AA142757),
XX chicken CHD-W (W refers to the W chromosome) gene (AA142758) and
XX and the great tit CHD-W gene (AA142759). Translated amino acid
XX sequences of this region are provided in AAM08146-49. The CHD-1A
XX (see also AA142751) and CHD-W (see also AA142754-55) genes determine
XX sex in birds and can be used to identify the sex of an embryo,
XX foetus etc. and to manipulate the sex of progeny.
XX
XX Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

alignment_scores:
Quality: 27.00 Length: 27
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment\_block:  
US-08-973-363-9 x AAT42758 ..  
Align seg 1/1 to: AAT42758 from: 1 to: 153

8 LysLeuProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrIleuIlely 24  
|||||  
52 AAGAAACCCAGCGCTAAGCAGTTACAGACCCGTCAGATTACCTCATTTAA 101  
|||||  
24 sLeuLeuAsnLysAspLeuAlaArgLysGlu 34  
|||||  
102 ATTACTGTAATTAAGACCTTCAGCAAGAAAGGA 132  
|||||

seq\_name: /SIDSI/gcgcdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV59280

seq\_documentation\_block:

ID AAV59280 standard: CDNA: 1311 BP.

AAV59280;

14-DEC-1998 (first entry)

DE Altered telomere repeat binding factor 1 gene.

KW ds; human; telomere repeat binding factor; A-TRF; dimerisation domain;

telomere; ageing; ataxia telangiectasia; Down's syndrome; tumour; viral.

OS Homo sapiens.

XX Synthetic.

Key Location/Qualifiers

FT 1..1311

FT /\*tag= a

FT /product= "A-TRF"

XX WO9836066-A1.

XX 20-AUG-1998.

XX PF 13-FEB-1998; 98WO-US02765.

XX PR 04-FEB-1998; 98US-0018628.

XX PR 13-FEB-1997; 97US-0800264.

XX PA (UYRQ ) UNIV ROCKEFELLER.

XX PI Bianchi A, De Lange T, Van Steensel B;

XX WPI: 1998-480769/41.

XX DR P-PSDB: AAV59280.

XX Nucleic acid encoding altered telomere repeat binding protein and  
XX related vectors - transformants, hetero-dimers and antibodies, used  
XX to inhibit shortening of telomerases caused by ageing or disease,  
XX also used to extend life of cells in culture

XX Claim 14; Page 110-111; 163pp; English.

XX The altered vertebrate telomere repeat binding protein (A-TRF) has a  
XX telomere repeat binding factor (TRF) dimerisation domain, and forms a  
XX hetero-dimer with TRF, preventing it from binding to the specified repeat  
XX sequence. A-TRF, optionally expressed by gene therapy, is used to  
XX inhibit shortening of telomeres associated with ageing (for cosmetic  
XX purposes) and disease, e.g. ataxia telangiectasia, Down's syndrome,  
XX atrophy of the skin, age-related macular degeneration, atherosclerosis,  
XX tumours and viral (including human immune deficiency virus) infection.  
XX Cells expressing A-TRF also have an increased life span in vitro, e.g.  
XX for expression of recombinant proteins or where intended for subsequent  
XX transplant or for testing, eliminating the need for transformation.

XX Sequence 1311 BP; 441 A; 257 C; 316 G; 297 T; 0 other;

alignment\_scores:  
Quality: 25.00 Length: 25  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-9 x AAV59280 ..

Align seg 1/1 to: AAV59280 from: 1 to: 1311

1 lLeuProAspAspProAspLysLysProGlnAlaLysGlnLeuGlnTh 17  
|||||  
1237 ATTCTTCAGATGATCCTGATTAAGAAACCAACGCAAAACGTTACAGAC 1286  
|||||  
17 rArgAlaAspTyrIleuIleLysIleu 25  
|||||  
1287 CCGTGCAGACTACTCTCATCAAACTA 1311  
|||||

seq\_name: /SIDSI/gcgcdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AAV42756

seq\_documentation\_block:

ID AAT42756 standard: DNA: 153 BP.

XX AAT42756;

XX 12-MAR-1997 (first entry)

XX DE Mouse CHD-1 gene (bases 3855-9777).

XX KW Bird; sex determination; chromodomain-Helicase-DNA binding 1;

XX CHD-1; CHD-W; W Chromosome; ss.

XX Mus sp.

Key Location/Qualifiers

FT 52..81

FT /\*tag= a

FT /note= "bases 52-81 are a repeat of bases 22-51  
and are ignored in the translated amino  
acid sequence given in Fig 3"

XX WO9639505-A1.

XX 12-DEC-1996.

XX PF 05-JUN-1996; 96WO-GB01341.

XX PR 06-JUN-1995; 95GB-0011439.

XX PA (ISIS-) ISIS INNOVATION LTD.

XX PI Griffiths R, Tiwari B;

XX WPI: 1997-043127/04.

XX DR P-PSDB: AAW08146.

XX Avian chromodomain-helicase-DNA binding genes determine sex in  
XX birds - used for sex determin. and to control sex of progeny  
XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
XX to portions of the chicken CHD-1A (A = Avian) gene (see also  
XX and AAT42757). chicken CHD-W (W refers to the W chromosome) gene  
XX (see also AAT42758) and the great tit CHD-W gene (see also AAT42759).  
XX Translated amino acid sequences of this region are provided in  
XX AAW08146-49. The CHD-1A (see also AAT42751) and CHD-W (see also  
XX AAT42754-55) genes determine sex in birds and can be used to identify  
XX the sex of an embryo, fetus etc. and to manipulate the sex of  
XX progeny.

XX Sequence 153 BP; 61 A; 37 C; 27 G; 28 T; 0 other;

## alignment\_scores:

Quality: 19.00 Length: 19  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-9 x AAT42756 ..

Align seg 1/1 to: AAT42756 from: 1 to: 153

8 LysLysProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLeuIlely 24  
|||||  
52 AAAAAAAAAACAGCAAGCAAGTACAGACCCGTCGACACTACCTCATCA 101

24 stleu 26  
|||||  
102 ACTACTT 108

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06443

## seq\_documentation\_block:

ID ABL06443 standard; cDNA: 6240 BP.

XX ABL06443:

XX AC ABL06443:

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13811.

XX KW Drosophila: developmental biology; cell signalling; insecticide;

XX KM pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN MO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001MO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR P-PSDB; ABB62340.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT interactions -

XX PS Claim 1; SEQ ID NO 13811; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB130511), expressed DNA sequences (AB57737-AB872072) and the encoded proteins

CC (AB57737-AB872072)  
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

CC CC Sequence 6240 BP; 1675 A; 1682 C; 1760 G; 1123 T; 0 other;

## alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-9 x ABL06443 ..

Align seg 1/1 to: ABL06443 from: 1 to: 6240

9 LysProGlnAlaLysGlnLeuGlnThrArgAla 19  
|||||  
4205 AAGCCCAAGCCAGCAGCTGCGACGCGTGC 4237

seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:ABL06442

## seq\_documentation\_block:

ID ABL06442 standard; cDNA: 9933 BP.

XX ABL06442:

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 13808.

XX KW Drosophila: developmental biology; cell signalling; insecticide;

XX KM pharmaceutical; gene; ss.

XX OS Drosophila melanogaster.

XX PN MO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001MO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE ) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR P-PSDB; ABB62339.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX PT interactions -

XX PS Claim 1; SEQ ID NO 13808; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (AB16176-AB130511), expressed DNA sequences (AB57737-AB872072) and the encoded proteins

CC (AB57737-AB872072)  
CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

CC CC Sequence 9933 BP; 2711 A; 2468 C; 2538 G; 2216 T; 0 other;

## alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-9 x ABL06442 ..

Align seg 1/1 to: ABL06442 from: 1 to: 9933  
9 LysProGlnAlaIysGlnLeuGlnThrArgAla 19  
|||||  
6745 AAGCCCAAGGCCAAGCAGCTGCACGCCGTGCC 6777  
seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseg-emb1/NA2001A.DAT:AAK88882  
seq\_documentation\_block:  
ID AAK88882 standard: cDNA, 421 BP.  
XX AAK88882;  
AC  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen coding sequence SEQ ID NO: 1198.  
XX  
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
digestive system disorder; Meckel's diverticulum; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20015314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01324.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.

```

XX 17-NOV-2000: 2000US-0249244.
XX 17-NOV-2000: 2000US-0249245.
XX 17-NOV-2000: 2000US-0249246.
XX 17-NOV-2000: 2000US-0249265.
XX 17-NOV-2000: 2000US-0249265.
XX 17-NOV-2000: 2000US-0249297.
XX 17-NOV-2000: 2000US-0249299.
XX 17-NOV-2000: 2000US-0249300.
XX 01-DEC-2000: 2000US-0250150.
XX 01-DEC-2000: 2000US-0250391.
XX 05-DEC-2000: 2000US-0251030.
XX 05-DEC-2000: 2000US-0251988.
XX 05-DEC-2000: 2000US-0256719.
XX 06-DEC-2000: 2000US-0251479.
XX 08-DEC-2000: 2000US-0251856.
XX 08-DEC-2000: 2000US-0251858.
XX 08-DEC-2000: 2000US-0251869.
XX 08-DEC-2000: 2000US-0251889.
XX 08-DEC-2000: 2000US-0251990.
XX 11-DEC-2000: 2000US-0254097.
XX 05-JAN-2001: 2001US-0239678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SW;
XX
XX WPI: 2001-502630/55.
XX
XX P-PSDB: AAM93109.
XX
XX
XX Polynucleotides encoding digestive system antigens, useful for
XX PT diagnosing, treating, preventing and/or pronozing disorders of the
XX PT digestive system, particularly cancer and cancer metastases -
XX
XX Claim 1: SEQ ID NO 1198; 986pp: English.
XX
XX
XX The present invention provides the protein and coding sequences of a
XX CC number of human digestive system antigens. These can be used in the
XX CC diagnosis, treatment and prevention of digestive system disorders,
XX CC including, treatment and prevention of digestive system disorders,
XX CC infections, appendicitis, Meckel's diverticulum, bacterial or parasitic
XX CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
XX CC ulcerative colitis. The present sequence is a cDNA encoding a digestive
XX CC system antigen of the invention.
XX
XX
XX Sequence 421 BP; 148 A; 78 C; 127 G; 61 T; 7 other;
XX
XX
XX Alignment scores:
XX
XX Quality: 10.00 Length: 10
XX Ratio: 1.000 Gaps: 0
XX Percent Similarity: 100.000 Percent Identity: 100.000
XX
XX alignment_block:
XX US-08-973-363-9 x AAK88882 ..
XX
XX Align seg 1/1 to: AAK88882 from: 1 to: 421
XX
XX 13 LysGlnLeuGlnThrArgAlaAspTyrLeu 22
XX ||||||||||||||||||||||||
XX 146 AAGCAGCTACGACGCCGAGCGATTACTTG 175
XX
XX seq_name: /SISL/gcgdata/hold-geneseq/genesegn-emb1/NA2001A.DAT:AA157603
XX
XX seq_documentation_block:
XX ID AA157603 standard; cDNA; 421 BP.
XX
XX AC AA157603;
XX
XX DT 19-OCT-2001 (first entry)
XX
XX DE Human colorectal cancer antigen cDNA SEQ ID NO: 67.
XX
XX KW Human; colorectal cancer; colorectal cancer antigen; gene therapy; ss.
XX
XX OS Homo sapiens.

```

XX	PN	W0200155350-AL.	
XX	PN	02-AUG-2001.	
XX	PF	17-JAN-2001.	2001NOV-0501350
XX	XX	31-JAN-2000.	2000OCT-0179065
PR	PR	04-FEB-2000.	2000OCT-0160628
PR	PR	24-FEB-2000.	2000OCT-0184664
PR	PR	02-MAR-2000.	2000OCT-0186350
PR	PR	16-MAR-2000.	2000OCT-0189674
PR	PR	17-MAR-2000.	2000OCT-0190876
PR	PR	18-APR-2000.	2000OCT-0198123
PR	PR	19-MAY-2000.	2000OCT-0205515
PR	PR	07-JUN-2000.	2000OCT-0209457
PR	PR	28-JUN-2000.	2000OCT-0214866
PR	PR	30-JUN-2000.	2000OCT-0215135
PR	PR	07-JUL-2000.	2000OCT-0216647
PR	PR	07-JUL-2000.	2000OCT-0216880
PR	PR	11-JUL-2000.	2000OCT-0217487
PR	PR	11-JUL-2000.	2000OCT-0217496
PR	PR	14-JUL-2000.	2000OCT-0218290
PR	PR	26-JUL-2000.	2000OCT-0220954
PR	PR	26-JUL-2000.	2000OCT-0220963
PR	PR	14-AUG-2000.	2000OCT-0224519
PR	PR	14-AUG-2000.	2000OCT-0224519
PR	PR	14-AUG-2000.	2000OCT-0225213
PR	PR	14-AUG-2000.	2000OCT-0225214
PR	PR	14-AUG-2000.	2000OCT-0225267
PR	PR	14-AUG-2000.	2000OCT-0225267
PR	PR	14-AUG-2000.	2000OCT-0225270
PR	PR	14-AUG-2000.	2000OCT-0225477
PR	PR	14-AUG-2000.	2000OCT-0225757
PR	PR	14-AUG-2000.	2000OCT-0225759
PR	PR	18-AUG-2000.	2000OCT-0226279
PR	PR	22-AUG-2000.	2000OCT-0226681
PR	PR	22-AUG-2000.	2000OCT-0226868
PR	PR	23-AUG-2000.	2000OCT-0227182
PR	PR	30-AUG-2000.	2000OCT-0228790
PR	PR	01-SEP-2000.	2000OCT-0228947
PR	PR	01-SEP-2000.	2000OCT-0229343
PR	PR	01-SEP-2000.	2000OCT-0229344
PR	PR	01-SEP-2000.	2000OCT-0229345
PR	PR	05-SEP-2000.	2000OCT-0229509
PR	PR	05-SEP-2000.	2000OCT-0229517
PR	PR	06-SEP-2000.	2000OCT-0230437
PR	PR	06-SEP-2000.	2000OCT-0230438
PR	PR	08-SEP-2000.	2000OCT-0231242
PR	PR	08-SEP-2000.	2000OCT-0231243
PR	PR	08-SEP-2000.	2000OCT-0231244
PR	PR	08-SEP-2000.	2000OCT-0231414
PR	PR	08-SEP-2000.	2000OCT-0232080
PR	PR	08-SEP-2000.	2000OCT-0232081
PR	PR	12-SEP-2000.	2000OCT-0231968
PR	PR	14-SEP-2000.	2000OCT-0232386
PR	PR	14-SEP-2000.	2000OCT-0232389
PR	PR	14-SEP-2000.	2000OCT-0232403
PR	PR	14-SEP-2000.	2000OCT-0232403
PR	PR	14-SEP-2000.	2000OCT-0233064
PR	PR	21-SEP-2000.	2000OCT-0234223
PR	PR	21-SEP-2000.	2000OCT-0234223
PR	PR	25-SEP-2000.	2000OCT-0234997
PR	PR	25-SEP-2000.	2000OCT-0235448
PR	PR	26-SEP-2000.	2000OCT-0235844

XX	27-SEP-2000;	2000US-0235836.	PR
XX	29-SEP-2000;	2000US-0236327.	PR
XX	29-SEP-2000;	2000US-0236357.	PR
XX	29-SEP-2000;	2000US-0236358.	PR
XX	29-SEP-2000;	2000US-0236359.	PR
XX	29-SEP-2000;	2000US-0236370.	PR
XX	02-OCT-2000;	2000US-0236802.	PR
XX	02-OCT-2000;	2000US-0237037.	PR
XX	02-OCT-2000;	2000US-0237038.	PR
XX	02-OCT-2000;	2000US-0237039.	PR
XX	02-OCT-2000;	2000US-0237040.	PR
XX	13-OCT-2000;	2000US-0239935.	PR
XX	13-OCT-2000;	2000US-0239937.	PR
XX	20-OCT-2000;	2000US-0240960.	PR
XX	20-OCT-2000;	2000US-0241221.	PR
XX	20-OCT-2000;	2000US-0241785.	PR
XX	20-OCT-2000;	2000US-0241786.	PR
XX	20-OCT-2000;	2000US-0241787.	PR
XX	20-OCT-2000;	2000US-0241808.	PR
XX	20-OCT-2000;	2000US-0241809.	PR
XX	20-OCT-2000;	2000US-0241826.	PR
XX	01-NOV-2000;	2000US-0244617.	PR
XX	08-NOV-2000;	2000US-0246474.	PR
XX	08-NOV-2000;	2000US-0246475.	PR
XX	08-NOV-2000;	2000US-0246476.	PR
XX	08-NOV-2000;	2000US-0246477.	PR
XX	08-NOV-2000;	2000US-0246478.	PR
XX	08-NOV-2000;	2000US-0246523.	PR
XX	08-NOV-2000;	2000US-0246524.	PR
XX	08-NOV-2000;	2000US-0246525.	PR
XX	08-NOV-2000;	2000US-0246526.	PR
XX	08-NOV-2000;	2000US-0246527.	PR
XX	08-NOV-2000;	2000US-0246528.	PR
XX	08-NOV-2000;	2000US-0246532.	PR
XX	08-NOV-2000;	2000US-0246609.	PR
XX	08-NOV-2000;	2000US-0246610.	PR
XX	08-NOV-2000;	2000US-0246611.	PR
XX	08-NOV-2000;	2000US-0246613.	PR
XX	17-NOV-2000;	2000US-0249207.	PR
XX	17-NOV-2000;	2000US-0249208.	PR
XX	17-NOV-2000;	2000US-0249209.	PR
XX	17-NOV-2000;	2000US-0249210.	PR
XX	17-NOV-2000;	2000US-0249211.	PR
XX	17-NOV-2000;	2000US-0249212.	PR
XX	17-NOV-2000;	2000US-0249213.	PR
XX	17-NOV-2000;	2000US-0249214.	PR
XX	17-NOV-2000;	2000US-0249215.	PR
XX	17-NOV-2000;	2000US-0249216.	PR
XX	17-NOV-2000;	2000US-0249217.	PR
XX	17-NOV-2000;	2000US-0249218.	PR
XX	17-NOV-2000;	2000US-0249244.	PR
XX	17-NOV-2000;	2000US-0249245.	PR
XX	17-NOV-2000;	2000US-0249264.	PR
XX	17-NOV-2000;	2000US-0249265.	PR
XX	17-NOV-2000;	2000US-0249297.	PR
XX	17-NOV-2000;	2000US-0249299.	PR
XX	17-NOV-2000;	2000US-0249300.	PR
XX	01-DEC-2000;	2000US-0250160.	PR
XX	01-DEC-2000;	2000US-0250391.	PR
XX	05-DEC-2000;	2000US-0251030.	PR
XX	05-DEC-2000;	2000US-0251030.	PR
XX	05-DEC-2000;	2000US-0251988.	PR
XX	05-DEC-2000;	2000US-0256719.	PR
XX	06-DEC-2000;	2000US-0251479.	PR
XX	08-DEC-2000;	2000US-0251856.	PR
XX	08-DEC-2000;	2000US-0251856.	PR
XX	08-DEC-2000;	2000US-0251869.	PR
XX	08-DEC-2000;	2000US-0251989.	PR
XX	08-DEC-2000;	2000US-0251990.	PR
XX	11-DEC-2000;	2000US-0254097.	PR
XX	05-JAN-2001;	2001US-0259678.	PR
XX	(HUMA- )	HUMAN GENOME SCI INC.	XX
XX			XX
PI	Rosen CA, Barash SC, Ruben SM;		
XX	WPI, 2001-457727/49.		
DR	P-PSDB; AAM38625.		
XX			
PT	Isolated polypeptide for treating, preventing and/or prognosing		
PT	disorders related to the colon and rectum including colorectal cancers		
PT	and also for testing and detection e.g. diagnosis -		
XX			
PS	Claim 1; SEQ ID NO: 67; 522pp + Sequence Listing; English.		
XX			
CC	The present invention provides the protein and coding sequences of a		
CC	number of colorectal cancer antigens. These are shown in		
CC	AA157547-AA157619 and AAM38569-AAM38641. These can be used in the		
CC	diagnosis, prevention and treatment of cancer of the colon and/or rectum		
CC	The present sequence is a colorectal cancer antigen coding sequence of		
CC	the invention.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC			



XX Claim 1; Columns 29-30; 30pp; English.  
XX  
XX The invention provides polynucleotides which encode novel protein  
CC kinase homologs expressed in various human cells and tissues. The  
CC present sequence represents the clone 23789 cDNA fragment derived  
CC from a human inflamed adenoid cDNA library. The cDNA encodes a  
CC protein kinase which shows homology to the CHKFRK chicken tyr kinase.  
CC Vectors and host cells can be used for recombinant production of the  
CC protein kinase homolog. The recombinant proteins may be used to raise  
CC antibodies for use as anti-kinase therapeutics. Oligonucleotides based  
CC on the polynucleotide sequences, i.e. probes and antisense constructs,  
CC the peptides and antibodies are claimed to be useful as tools for  
CC studying signalling cascades in cells and proteins, and for identifying  
CC inhibitors (drugs) to treat diseases and inflammatory conditions  
CC associated with abnormal kinase expression. Diseases that are claimed  
CC to be treatable include human X-linked agammaglobulinemia,  
CC nonspherocytic hemolytic anaemia, artherosclerosis, carcinomas,  
CC diabetes, glaucoma, restenosis, cholera-based septic shock, etc.  
XX  
SQ Sequence 252 BP; 63 A; 70 C; 72 G; 47 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-9 x AAV33521 ..  
  
Align seg 1/1 to: AAV33521 from: 1 to: 252  
  
24 TyrosineLeuAnlysAspLeuAla 31  
|||||  
100 AACTGCTCAACAAAGACTGCGA 123  
  
seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA61903  
  
seq\_documentation\_block:  
ID ABA61903 standard; DNA: 544 BP.  
XX  
XX ABA61903;  
XX  
XX 01-FEB-2002 (first entry)  
XX  
XX Human foetal liver single exon nucleic acid probe #10208.  
XX  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157277-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-0500669.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX  
XX 26-MAY-2000; 2000US-0207456.  
XX  
XX 30-JUN-2000; 2000US-0608408.  
XX  
XX 03-AUG-2000; 2000US-0632366.  
XX  
XX 21-SEP-2000; 2000US-0234687.  
XX  
XX 27-SEP-2000; 2000US-0236359.  
XX  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -  
XX  
XX Claim 1; SEQ ID NO 10208; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published-pct-sequences.  
XX  
SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-9 x ABA61903 ..  
  
Align seg 1/1 to: ABA61903 from: 1 to: 544  
  
21 TyrosineLeuAnlys 28  
|||||  
243 TACCTGATTAAGTATTAATTAAG 266  
  
seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK10214  
  
seq\_documentation\_block:  
ID AAK10214 standard; DNA: 544 BP.  
XX  
XX AAK10214;  
XX  
XX 05-NOV-2001 (first entry)  
XX  
XX Human brain expressed single exon probe SEQ ID NO: 10205.  
XX  
XX Human; brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200157275-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-0500667.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX  
XX 26-MAY-2000; 2000US-0207456.  
XX  
XX 30-JUN-2000; 2000US-0608408.  
XX  
XX 03-AUG-2000; 2000US-0632366.  
XX  
XX 21-SEP-2000; 2000US-0234687.  
XX  
XX 27-SEP-2000; 2000US-0236359.  
XX  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX  
XX Example 4; SEQ ID NO: 10205; 650pp + Sequence listing; English.  
XX

CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is one of the probes of the  
 CC invention.  
 CC  
 SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-9 x AAK10214 ..  
 Align seg 1/1 to: AAK10214 from: 1 to: 544  
 21 TyrleuilelyslleuAsnlys 28  
 ||||||||||||||||||||  
 243 TACCTGATTAAGTATTATAATAAG 266  
 seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAK36112  
 seq\_documentation\_block:  
 ID AAK36112 standard: DNA; 544 BP.  
 AC AAK36112;  
 DT 06-NOV-2001 (first entry)  
 DE Human bone marrow expressed single exon probe SEQ ID NO: 10669.  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.  
 OS Homo sapiens.  
 PN WO200157276-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001; 2001WO-US00668.  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI: 2001-488900/53.  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 PS Example 4: SEQ ID NO: 10669; 658bp + Sequence Listing; English.  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
 CC the probes of the invention.  
 CC


SQ Sequence 544 BP; 144 A; 128 C; 88 G; 184 T; 0 other;  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-9 x AAK36112 ..  
 Align seg 1/1 to: AAK36112 from: 1 to: 544  
 21 TyrleuilelyslleuAsnlys 28  
 ||||||||||||||||||||  
 243 TACCTGATTAAGTATTATAATAAG 266

Mon Aug 5 11:52:02 2002

us-08-973-363-9.olip2n.rng

---

Page 11



**THIS PAGE BLANK (USPTO)**



DEFINITION AL644594 XGC-egg silurana tropicalis cDNA clone l1e1d12 5', mRNA sequence.  
ACCESSION AL644594  
VERSION AL644594.1 GI:16796719  
KEYWORDS EST.  
SOURCE western clawed frog.  
ORGANISM silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Silurana.  
REFERENCE 1 (bases 1 to 645)  
Huckle,E., Taylor,R., Ashurst,J.L., Zorn,A.M. and Rogers,J.  
Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
JOURNAL Unpublished (2001)  
CONTACT: Huckle E  
Sanger Centre  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: tropesanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: l1e1d12.pic  
Sequencing primer: p1c  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
FEATURES  
source  
1..645  
/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="l1e1d12"  
/clone\_lib="XGC-egg"  
/dev\_stage="egg"  
/lab\_host="Escherichia coli XL1-blue"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dt primed from 5ug of poly A+ RNA from egg. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end"  
BASE COUNT 222 a 125 c 156 g 141 t 1 others  
ORIGIN  
alignment\_scores:  
Quality: 30.00 Length: 30  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-9 x AL644594 ..  
Align seg 1/1 to: AL644594 from: 1 to: 645  
1 l1e1euProaspPProaspPlyslsProGlnAlaLysGlnLeuGlnth 17  
|||||  
456 ATTTCACCATGATCCAGATAGAACCCCAAGCCAGCAGCAGAC 505  
17 rArgAlaAspTyrLeuIleLysLeuLeuAsnLysAspLeu 30  
|||||  
506 CAGAGCTGACTACCTCATTAACCTTCATTAAGATCTG 545  
seq\_name: gb\_est1:AL601246  
seq\_documentation\_block:  
LOCUS AL601246 430 bp mRNA linear EST 14-AUG-2001  
DEFINITION DKFZP313J1040\_r1 313 (synonym: hlcc2) Homo sapiens cDNA clone  
DKFZP313J1040\_5', mRNA sequence.  
ACCESSION AL601246  
VERSION AL601246.1 GI:15164752  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 430)  
Bloeker,H., Boecher,M., Brandt,P., Mewes,W., Well,B. and Wiemann  
,S.

TITLE EST (Bloeker,H., Boecher,M., Brandt,P., Mewes,H.W., Well,B. and Wiemann,S.)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Bloeker H  
MIPS  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 5' sequence of the clone insert  
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
Sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.  
No sl sequence available.  
This clone (DKFZP313J1040) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY. Email: clone@rzd.de.  
FEATURES  
source  
1..430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZP313J1040"  
/clone\_lib="313 (synonym: hlcc2)"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pTriplex2; Site\_1: SfiIA; Site\_2: SfiIB; cDNA-collection"  
BASE COUNT 163 a 81 c 86 g 100 t  
ORIGIN  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-9 x AL601246 ..  
Align seg 1/1 to: AL601246 from: 1 to: 430  
1 l1e1euProaspPProaspPlyslsProGlnAlaLysGlnLeuGlnth 17  
|||||  
218 ATTCTCCAGATGATCCGATTAATAAACCAACCAAGCAGTTCAGAC 267  
17 rArgAlaAspTyrLeuIleLysLeuLeu 26  
|||||  
268 CCGTGCAGACTACCTCATCAATTAATTCTT 295  
seq\_name: gb\_est1:A1890775  
seq\_documentation\_block:  
LOCUS A1890775 547 bp mRNA linear EST 07-MAR-2000  
DEFINITION wm95f11.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:2443725 3  
similar to SW:CH1\_HUMAN O14646 CHROMODOMAIN-HELICASE-DNA-BINDING  
PROTEIN 1; , mRNA sequence.  
ACCESSION A1890775  
VERSION A1890775.1 GI:5595939  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 547)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.



1 l1leuPProAspAspProAspLysProGlnAlaLysGlnLeuGlnrth 17  
|||||  
489 ATCTTCAGATGATCCTGATTAACCAACCAACAGTTACGAC 538  
17 l7aRGAlaAspTyrLeuIleLysLeu 26  
|||||  
539 CCGTGCAGACTACCTCATCAACTACTT 566  
seq\_name: gb\_est1:BB461065  
seq\_documentation\_block:  
LOCUS BB461065 660 bp mRNA linear EST 25-OCT-2001  
DEFINITION BB461065 RIKEN full-length enriched, 12 days embryo spinal ganglion  
Mus musculus cDNA clone D130070B13 3' similar to l10410 Mouse  
DNA-binding protein (CHD-1) mRNA, mRNA sequence.  
ACCESSION BB461065  
VERSION BB461065.2 GI:16426612  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 660)  
Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,  
Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,  
M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,  
D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,  
Tagami,M., Tagawa,A., Takahashi,F., Takeba,Y., Tanaka,T., Toya,T.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)  
Unpublished (2001)  
TITLE On Jul 21, 2000 this sequence version replaced gi:9356558.  
JOURNAL Contact: Yoshhide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp  
URL: http://genome.gsc.riken.go.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh  
M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura  
S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and  
Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara  
Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanka,I., Aizawa  
K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and  
Hayashizaki,Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
FEATURES  
Source location/Qualifiers  
1. .660

/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D130070B13"  
/clone\_lib="RIKEN full-length enriched, 12 days embryo  
spinal ganglion"  
/tissue\_type="spinal ganglion"  
/dev\_stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site 1: Saito, Site 2: BamHI, cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5',  
GAGAGAGAGAGCGCGCCGACACTCGAGTTTCTTTTCTTTTCTTTVN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence [5',  
GAGAGAGAGATTCGAGTTATTTAAATTAATCCCCCCCCC 3']. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 211 a 130 c 168 g 151 t  
ORIGIN  
alignment\_scores:  
Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-9 x BB461065 ..  
Align seg 1/1 to: BB461065 from: 1 to: 660  
1 l1leuPProAspAspProAspLysProGlnAlaLysGlnLeuGlnrth 17  
|||||  
481 ATCTTCAGATGATCCTGATTAACCAACCAACAGTTACGAC 530  
17 l7aRGAlaAspTyrLeuIleLysLeu 26  
|||||  
531 CCGTGCAGACTACCTCATCAACTACTT 558  
seq\_name: gb\_est1:AU125712  
seq\_documentation\_block:  
LOCUS AU125712 866 bp mRNA linear EST 23-OCT-2000  
DEFINITION AU125712 NT2RM4 Homo sapiens cDNA clone NT2RM4002061 5', mRNA  
sequence.  
ACCESSION AU125712  
VERSION AU125712.1 GI:10950428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 866)  
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,  
Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and  
Isigai,T.  
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,  
Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki  
Y., Sugano,S., Isigai,T.)  
Unpublished (2000)  
COMMENT Contact: Takao Isigai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp



HRI human cDNA project: 5'-6'-3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

## FEATURES

source

Location/Qualifiers

1. 866

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="NT2RM4002061"

/clone\_lib="NT2RM4"

/cell\_type="teratocarcinoma"

/cell\_line="NT2"

/note="Vector: pME18SFLJ; mRNA from uninduced NT2 neuronal precursor cells"

BASE COUNT 312 a 149 c 196 g 207 t 2 others

ORIGIN

## alignment\_scores:

Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AU125712

Align seg 1/1 to: AU125712 from: 1 to: 866

1 11leuProaspAspProaspLysLysProGlnAlaLysGlnLeuGlnh 17  
|||||  
450 ATCTCTCCAGATGATCCGATATAAACACAGCAAAAGAGTTGCAGAC 499  
|||||  
17 TARGAlaspTyrLeuLeuLysLeuLeu 26  
|||||  
500 CCGTCAGACATCACTCATCAATTAATCTT 527

seq\_name: gb\_est2:BE895133

seq\_documentation\_block:

LOCUS BE895133 1028 bp mRNA linear EST 20-OCT-2000  
DEFINITION 601436060P1 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:3921087 5',  
mRNA sequence.  
ACCESSION BE895133  
VERSION BE895133.1 GI:10358221  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1028)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DRP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LHM9753 row: h column: 16  
High quality sequence stop: 488.

## FEATURES

source

Location/Qualifiers

1. 1028

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:3921087"

/clone\_lib="NIH\_MGC\_72"

/tissue\_type="melanocytic melanoma"

/lab\_host="DH10B (Phage-resistant)"

/note="Organ: skin; Vector: PCMV-SPORT6; Site: 1; Nct1;

Site: 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 2 kb. Library constructed by Life Technologies."

## BASE COUNT

387 a 205 c 238 g 198 t

ORIGIN

## alignment\_scores:

Quality: 26.00 Length: 26  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x BE895133

Align seg 1/1 to: BE895133 from: 1 to: 1028

1 11leuProaspAspProaspLysLysProGlnAlaLysGlnLeuGlnh 17  
|||||  
131 ATCTCTCCAGATGATCCGATATAAACACAGCAAAAGAGTTGCAGAC 180  
|||||  
17 TARGAlaspTyrLeuLeuLysLeuLeu 26  
|||||  
181 CCGTCAGACATCACTCATCAATTAATCTT 208

seq\_name: gb\_est1:AW96787

seq\_documentation\_block:

LOCUS AW96787 337 bp mRNA linear EST 05-JUN-2000  
DEFINITION QV3-BN0047-230200-102-d03 BN0047 Homo sapiens CDNA, mRNA sequence.  
ACCESSION AW96787  
VERSION AW96787.1 GI:8257021  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 337)  
Dias Neto, E., Garcia Correa, R., Vertovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., de Silva, W. Jr., Zagzo, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Makukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20020663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=62-QV3-BN0047-230  
200-102-d03&f3=2000-02-23&t4=1)  
Seq primer: puc18 forward  
High quality sequence stop: 2  
High quality sequence start: 337.

## FEATURES

source

Location/Qualifiers

1. 337

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="BN0047"

/dev\_stage="Adult"

/note="Organ: breast, normal; Vector: puc18; Site: 1; SmaI;  
Site: 2: SmaI; A mini-library was made by cloning products  
derived from ONESTES PCR (U.S. Letters Patent application  
NO. 196,716 - Ludwig Institute for Cancer Research)

Profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 83 a 59 c 72 g 123 t

ORIGIN

alignment\_scores:

Quality:	23.00	Length:	23
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment\_block:

US-08-973-363-9 x AW996787/rev ..

Align seg 1/1 to reverse of: AW996787 from: 1 to: 337

```
1 ltleuProaspaspProaspLysProglAlaLysGlnLeuGlnth 17
|||||
105 ATTCTTCACAGATGATCCGATATAAACACAAAGCAAGATTGACGAC 56
17 rArgAlaAspTyrLeu 23
|||||
55 CCGTGACAGACTATCTCATC 37
```

seq\_name: gb\_est1:BB830730

seq\_documentation\_block:

LOCUS BB830730 438 bp mRNA linear EST 19-NOV-2001

DEFINITION BB830730 RIKEN full-length enriched, mammary gland RCB-0527 Jyg-MC(B) cDNA Mus musculus cDNA clone G930013K04 3', mRNA sequence.

ACCESSION BB830730

VERSION BB830730.1 GI:17008973

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunaseki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

Riken Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222 Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp

URL: http://genome-gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (<http://genome-gsc.riken.go.jp>) for further details.

FEATURES

source location/Qualifiers

1..438 /organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="G930013K04"

/clone\_1ib="RIKEN full-length enriched, mammary gland RCB-0527 Jyg-MC(B) cDNA"

/tissue\_type="mammary gland"

/cell\_line="RCB-0527 Jyg-MC(B)"

BASE COUNT 138 a 96 c 108 g 96 t

ORIGIN

alignment\_scores:

Quality:	22.00	Length:	22
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment\_block:

US-08-973-363-9 x BB830730 ..

Align seg 1/1 to: BB830730 from: 1 to: 438

```
1 ltleuProaspaspProaspLysProglAlaLysGlnLeuGlnth 17
|||||
300 ATTCTTCACAGATGATCCGATATAAACACAAAGCAAGATTGACGAC 349
17 rArgAlaAspTyrLeu 22
|||||
350 CCGTGACAGACTATCTCATC 365
```

seq\_name: gb\_est1:BB834922

seq\_documentation\_block:

LOCUS BB834922 446 bp mRNA linear EST 19-NOV-2001

DEFINITION BB834922 RIKEN full-length enriched, mammary gland RCB-0527 Jyg-MC(B) cDNA Mus musculus cDNA clone G930033J21 3', mRNA sequence.

ACCESSION BB834922

VERSION BB834922.1 GI:17013165

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,U., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunaseki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

Riken Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222 Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp

URL: http://genome-gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)

wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Iwata, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.

Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.

# FEATURES

source  
1. 446  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="G930033j21"  
/clone\_1lb="RIKEN full-length enriched, mammary gland  
RCB-0527 Jy9-MC(B) cDNA"  
/tissue\_type="mammary gland"  
/cell\_line="RCB-0527 Jy9-MC(B)"  
BASE COUNT 139 a 99 c 108 g 100 t  
ORIGIN

alignment\_scores:  
Quality: 22.00 Length: 22  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-9 x BB834922 ..  
Align seg 1/1 to: BB834922 from: 1 to: 446

1 |||leupProaSpaSPProaSPlySPlySPProGlnAlaLysGlnLeuGlnTh 17  
|||||  
309 ATTCTCCAGATGATCTCATTAACCAACCAACCAACGACTACAGAC 358  
17 rArgAlaAspTyrLeu 22  
|||||  
359 CCGTCAGACACTACCTC 374

seq\_name: gb\_est2:BF239967

seq\_documentation\_block:

LOCUS BF239967 821 bp mRNA linear EST 14-NOV-2000  
DEFINITION 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129.5,  
mRNA sequence.

ACCESSION BF239967  
VERSION BF239967.1 GI:11153890  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 821)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

JOURNAL CDNA Library Preparation: CLONTECH Laboratories, Inc.  
COMMENT CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMN at: <http://image.llnl.gov>  
Plate: LLCMI03 row: k column: 18  
High quality sequence stop: 562.

# FEATURES

source  
1. 821  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4133129"  
/clone\_1lb="NIH\_MGC\_54"  
/tissue\_type="from chronic myelogenous leukemia"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: bone marrow; Vector: pDNR-LIB (Clontech); Site\_1: Still (ggccgctggcc); Site\_2: Still (ggccatlaagcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCGCCGCGCGCATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."  
BASE COUNT 345 a 149 c 190 g 137 t  
ORIGIN

alignment\_scores:  
Quality: 21.00 Length: 21  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-9 x BF239967 ..  
Align seg 1/1 to: BF239967 from: 1 to: 821

6 ProaSPlySPlySPProGlnAlaLysGlnLeuGlnThrArgAlaAspTyrLe 22  
|||||  
3 CCGGATTAACCAACCAACCAACCAACGACTACAGACTACCT 52  
22 uillelSPleu 26  
|||||  
53 CATCAAAATTACTT 65

seq\_name: gb\_est1:AW97058

seq\_documentation\_block:

LOCUS AW97058 686 bp mRNA linear EST 05-JUN-2000  
DEFINITION OV3-BN0047-150400-152-c03 BN0047 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AW97058  
VERSION AW97058.1 GI:8257292  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 686)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Magal, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-BN0047-150>)  
400-152-c036t3=2000-04-15&t=1)  
Seq primer: puc 18 forward  
High quality sequence start: 19  
High quality sequence stop: 678.

FEATURES  
source  
1..686  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="BN0047"  
/dev\_stage="Adult"

/note="Organ: breast\_normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT  
165 a 154 c 126 g 241 t

ORIGIN

alignment\_scores:  
Quality: 18.00 Length: 18  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x AM997058/rev ..

Align seg 1/1 to reverse of: AM997058 from: 1 to: 686

104 AAACCAACAGCAACAGTTCAGACCCGTCGACACTACCTCAATTT 55  
9 LysPProGlnAlaLysGlnLeuGlnThrArgAlaLysPtyrLeuIleLysLe 25  
|||||  
104 AAACCAACAGCAACAGTTCAGACCCGTCGACACTACCTCAATTT 55

25 uleu 26  
|||||  
54 ACTT 51

seq\_name: gb\_gss:CNS04DVG

seq\_documentation\_block:  
LOCUS CNS04DVG 856 bp DNA linear GSS 21-MAY-2000  
DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone  
103608 of library G from Tetradon nigroviridis, genomic survey  
sequence.  
AL286261  
VERSION AL286261.1 GI:8024707  
KEYWORDS GSS: genome survey sequence.  
SOURCE Tetradon nigroviridis.  
ORGANISM Tetradon nigroviridis.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.  
1 (bases 1 to 856)  
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,  
Bonneau,L., Billault,A., Quetier,F., Sautin,W., Bernot,A. and  
Weissenbach,J.  
Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetradon nigroviridis  
Unpublished  
2 (bases 1 to 856)  
Roest-Crolius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brothier,P., Quetier,F.,

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Saurin,W. and Weissenbach,J.  
Human gene number estimate provided by genome wide analysis using  
Tetradon nigroviridis DNA sequence  
Unpublished  
3 (bases 1 to 856)  
Genoscope.  
Direct Submission  
Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases  
This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetradon nigroviridis  
genome. For more information, please take a look at  
<http://www.genoscope.cns.fr/Tetradon>.

FEATURES  
source  
1..856  
Location/Qualifiers  
/organism="Tetradon nigroviridis"  
/db\_xref="taxon:99883"  
/clone\_lib="103K08"  
/clone\_lib="G"  
/note="Genoscope sequence ID : C0BG103BF04LPI-end : T7"

BASE COUNT  
84 a 308 c 252 g 206 t 6 others

ORIGIN

alignment\_scores:  
Quality: 18.00 Length: 18  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-9 x CNS04DVG/rev ..

Align seg 1/1 to reverse of: CNS04DVG from: 1 to: 856

2 LeuProAspAspProAspLysProGlnAlaLysGlnLeuGlnThrArg 18  
|||||  
674 CTGCGGATGATCTCTGATTAAGAGCCTCAGGCCAACAAGTTACAGCACG 625  
18 gAla 19  
|||||  
624 AGCC 621

seq\_name: gb\_est2:BF881342

seq\_documentation\_block:  
LOCUS BF881342 129 bp mRNA linear EST 17-JAN-2001  
DEFINITION PM0-ET0208-031200-001-f12 ET0208 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF881342  
VERSION BF881342.1 GI:12271468  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 129)  
Dias,Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Project: This entry can be seen in the following URL:  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l1=PM0&t2=PM0-ET0208-031200-001-f12&t3=2000-12-03&t4=1)

Seq primer: puc 18 forward  
 High quality sequence start: 22  
 High quality sequence stop: 128.  
 Location/Qualifiers

FEATURES

Source

1. 129  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="ET0208"  
 /dev\_stage="Adult"  
 /note="Organ: lung\_tumor; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the pUC 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."

BASE COUNT 21 a 48 c 23 g 37 t  
 ORIGIN

alignment\_scores:  
 Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-9 x BF881342/rev ..

Align seg 1/1 to reverse of: BF881342 from: 1 to: 129

13 LysGlnLeuGlnThrArgAlaSPtyrLeu 22  
 ||||||||||||||||||||||||||||  
 108 AAGCAGCTACAGACCGACGCGATTACTTG 79

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 01:43:18 ; Search time 10310.5 Seconds

(Without alignments)  
13411.753 Million cell updates/sec

Title: US-08-973-363-10

Perfect score: 6608

Sequence: 1 CGGGCTGGCGGCGAAGGCGC.....AACTAAACCAAAAAACC 6608

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenPdb1:\*  
1: gb\_ba:\*  
2: gb\_hkg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hkg\_hum:\*  
31: em\_hkg\_inv:\*  
32: em\_hkg\_other:\*  
33: em\_hkg\_inv:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	6608	100.0	6608	6	A58691	A58691 Sequence 10
2	4341	65.7	6872	5 <th>AF004397</th> <th>AF004397 Gallus ga</th>	AF004397	AF004397 Gallus ga
3	1410	21.3	2292	5 <th>D14316</th> <th>D14316 delta-cryst</th>	D14316	D14316 delta-cryst
4	155	2.3	345	5 <th>AF006659</th> <th>AF006659 Gallus ga</th>	AF006659	AF006659 Gallus ga
5	111	1.7	111	6 <th>A58713</th> <th>A58713 Sequence 32</th>	A58713	A58713 Sequence 32
6	111	1.7	2754	5 <th>AF181828</th> <th>AF181828 Nymphicus</th>	AF181828	AF181828 Nymphicus
7	103	1.6	153	6 <th>A58684</th> <th>A58684 Sequence 3</th>	A58684	A58684 Sequence 3
8	83	1.3	2754	5 <th>AF181824</th> <th>AF181824 Aegolius</th>	AF181824	AF181824 Aegolius
9	82	1.2	5947	9 <th>AF006513</th> <th>AF006513 Homo sapi</th>	AF006513	AF006513 Homo sapi
10	82	1.2	101220	9 <th>AC092372</th> <th>AC092372 Homo sapi</th>	AC092372	AC092372 Homo sapi
11	82	1.2	134365	2 <th>AC012624</th> <th>AC012624 Homo sapi</th>	AC012624	AC012624 Homo sapi
12	82	1.2	145659	2 <th>AC008531</th> <th>AC008531 Homo sapi</th>	AC008531	AC008531 Homo sapi
13	82	1.2	145659	2 <th>AC008531</th> <th>AC008531 Homo sapi</th>	AC008531	AC008531 Homo sapi
14	82	1.2	219228	2 <th>AC022121</th> <th>AC022121 Homo sapi</th>	AC022121	AC022121 Homo sapi
15	81	1.2	2754	5 <th>AF181827</th> <th>AF181827 Nymphicus</th>	AF181827	AF181827 Nymphicus
16	74	1.1	422	11 <th>GF3101</th> <th>GF3101 human STS W</th>	GF3101	GF3101 human STS W
17	74	1.1	2943	5 <th>AF181826</th> <th>AF181826 Gallus ga</th>	AF181826	AF181826 Gallus ga
18	72	1.1	71	1.1 <th>AF181825</th> <th>AF181825 Aegolius</th>	AF181825	AF181825 Aegolius
19	71	1.1	315	5 <th>AF288500</th> <th>AF288500 Phasianus</th>	AF288500	AF288500 Phasianus
20	67	1.0	67	5 <th>AF462045</th> <th>AF462045 Oreophaps</th>	AF462045	AF462045 Oreophaps
21	66	1.0	68827	2 <th>AC102745</th> <th>AC102745 Mus muscu</th>	AC102745	AC102745 Mus muscu
22	66	0.9	428	5 <th>AF138966</th> <th>AF138966 Centrocerc</th>	AF138966	AF138966 Centrocerc
23	59	0.9	309	5 <th>AF288503</th> <th>AF288503 Anas plat</th>	AF288503	AF288503 Anas plat
24	59	0.9	433	5 <th>AF138969</th> <th>AF138969 Chen caer</th>	AF138969	AF138969 Chen caer
25	59	0.9	439	5 <th>AF138968</th> <th>AF138968 Cygnus co</th>	AF138968	AF138968 Cygnus co
26	59	0.9	1316	6 <th>A58696</th> <th>A58696 Sequence 15</th>	A58696	A58696 Sequence 15
27	54	0.8	635	5 <th>AF059276</th> <th>AF059276 Struthio</th>	AF059276	AF059276 Struthio
28	50	0.8	309	5 <th>AF288501</th> <th>AF288501 Acryllium</th>	AF288501	AF288501 Acryllium
29	47	0.7	153	6 <th>A58685</th> <th>A58685 Sequence 4</th>	A58685	A58685 Sequence 4
30	47	0.7	309	5 <th>AF060700</th> <th>AF060700 Struthio</th>	AF060700	AF060700 Struthio
31	43	0.7	231	5 <th>AF276764</th> <th>AF276764 Ara marac</th>	AF276764	AF276764 Ara marac
32	43	0.7	231	5 <th>AF276766</th> <th>AF276766 Ara marac</th>	AF276766	AF276766 Ara marac
33	43	0.7	231	5 <th>AF276769</th> <th>AF276769 Bynopsit</th>	AF276769	AF276769 Bynopsit
34	42	0.6	325	5 <th>AF288499</th> <th>AF288499 Bonasa um</th>	AF288499	AF288499 Bonasa um
35	42	0.6	348	5 <th>AF060702</th> <th>AF060702 Struthio</th>	AF060702	AF060702 Struthio
36	42	0.6	143079	2 <th>AC021449</th> <th>AC021449 Homo sapi</th>	AC021449	AC021449 Homo sapi
37	42	0.6	193446	2 <th>AC091946</th> <th>AC091946 Homo sapi</th>	AC091946	AC091946 Homo sapi
38	42	0.6	193443	2 <th>AC092478</th> <th>AC092478 Homo sapi</th>	AC092478	AC092478 Homo sapi
39	42	0.6	193443	2 <th>AC092478</th> <th>AC092478 Homo sapi</th>	AC092478	AC092478 Homo sapi
40	41	0.6	276181	2 <th>AC092382</th> <th>AC092382 Homo sapi</th>	AC092382	AC092382 Homo sapi
41	41	0.6	435	3 <th>AF138964</th> <th>AF138964 Cygnus co</th>	AF138964	AF138964 Cygnus co
42	41	0.6	615	3 <th>EC112841</th> <th>EC112841 Rhabdites ci</th>	EC112841	EC112841 Rhabdites ci
43	41	0.6	615	3 <th>FAT12839</th> <th>FAT12839 Rhabdites ci</th>	FAT12839	FAT12839 Rhabdites ci
44	40	0.6	5349	10 <th>MOSCHDIX</th> <th>MOSCHDIX Mouse DNA-B</th>	MOSCHDIX	MOSCHDIX Mouse DNA-B
45	40	0.6	455	5 <th>AF138965</th> <th>AF138965 Chen caer</th>	AF138965	AF138965 Chen caer

## ALIGNMENTS

RESULT	1	LOCUS	6608 bp	DNA	linear	PAT 06-MAR-1998
A58691	A58691	Sequence 10 from Patent WO9639505.				
DEFINITION	A58691					
ACCESSION	A58691					
VERSION	A58691.1	GI:3714250				
KEYWORDS						
SOURCE	unidentified.					
ORGANISM	unclassified.					
REFERENCE	1 (bases 1 to 6608)					
AUTHORS	Griffiths, R. and Twari, B.					
TITLE	AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN					
JOURNAL	BIRDS					
COMMENT	PATENT: WO 9639505-A 10 12-DEC-1996;					
FEATURES	ISIS INNOVATION (GB)					
Other publication AU 5906996 961224.						
Location/Qualifiers						
source	1. 6608					
	/organism="unidentified"					
	/db_xref="taxon:32644"					
BASE COUNT	2289 a	1207 c	1459 g	1653 t		
ORIGIN						

Query Match 100.0%; Score 6608; DB 6; Length 6608;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 6608; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGGGCTGGGGGCGAGAGCGACCCGCGGCGACGCGAGGCTGGGGGGAAGGCTGGC 60  
Db 1 CGGGCTGGGGGCGAGAGCGACCCGCGGCGACGCGAGGCTGGGGGGAAGGCTGGC 60

QY 61 CGGGCGAGCCGGAGCGACGAGGTATTTGGGCAAAAATCTTGCCATCTGTAGAGATAG 120  
Db 61 CGGGCGAGCCGGAGCGACGAGGTATTTGGGCAAAAATCTTGCCATCTGTAGAGATAG 120

QY 121 CAAGTCAAAAGCATTACTTCGAAAAATAGGAGTACGAGAAAGGGGTTCTTGACCTAC 180  
Db 121 CAAGTCAAAAGCATTACTTCGAAAAATAGGAGTACGAGAAAGGGGTTCTTGACCTAC 180

QY 181 ACCTTGTAACTGAGTGGAGCTTTCTTTTAATCTTAACTTAACTTAACTTAACTTAACT 240  
Db 181 ACCTTGTAACTGAGTGGAGCTTTCTTTTAATCTTAACTTAACTTAACTTAACTTAACT 240

QY 241 GTGATGAAGAAGTGTAAAGAAACAGAGTGAAGTCAAGCATCAGATCAGATGATCTGT 300  
Db 241 GTGATGAAGAAGTGTAAAGAAACAGAGTGAAGTCAAGCATCAGATGATGATCTGT 300

QY 301 GGTAGCTTTCAGGTTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 360  
Db 301 GGTAGCTTTCAGGTTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 360

QY 361 GCCAGTCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 420  
Db 361 GCCAGTCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 420

QY 421 AGTGTACACATCTAGAT 480  
Db 421 AGTGTACACATCTAGAT 480

QY 481 CTGAGTTTGGAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 540  
Db 481 CTGAGTTTGGAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 540

QY 541 AGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCA 600  
Db 541 AGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCA 600

QY 601 GTAGTGAAGATTTGCGCGATGATCTGTCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 660  
Db 601 GTAGTGAAGATTTGCGCGATGATCTGTCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 660

QY 661 AAGACTGGCAAAATGTCAAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 720  
Db 661 AAGACTGGCAAAATGTCAAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTCAGGTC 720

QY 721 CGGAAAGTTGGGGATTAAGAGAGTGTGAAGAAAGTGAATCTGACTTAAGAGCAAAAAACA 780  
Db 721 CGGAAAGTTGGGGATTAAGAGAGTGTGAAGAAAGTGAATCTGACTTAAGAGCAAAAAACA 780

QY 781 AAGTCAAAAGCCGTTAACTTCAAGCAGATTAAGCAAAAGTGGGAAAAAGAGAGCAG 840  
Db 781 AAGTCAAAAGCCGTTAACTTCAAGCAGATTAAGCAAAAGTGGGAAAAAGAGAGCAG 840

QY 841 GACGAGAAAGAGCAACTTGATTCATCAGAGAGAGAGAGAGAGAGATGAAGATTATG 900  
Db 841 GACGAGAAAGAGCAACTTGATTCATCAGAGAGAGAGAGAGAGAGATGAAGATTATG 900

QY 901 ATAGAGAGAGATCTGCTGCGCAGGCAACAGTGAATGTATGTACAAAGAGCTGAAGAA 960  
Db 901 ATAGAGAGAGATCTGCTGCGCAGGCAACAGTGAATGTATGTACAAAGAGCTGAAGAA 960

QY 961 CCAAGAGAGATTTGATGATTTGCTGGAAGTTTGTGAGAGAGATGCCACAGACTGAAG 1020  
Db 961 CCAAGAGAGATTTGATGATTTGCTGGAAGTTTGTGAGAGAGATGCCACAGACTGAAG 1020

QY 1021 AAGATGATTTGAACATATAGAGAGATTATGACAGCTGCAATTTGGCCGGAAGAGAGCCA 1080  
Db 1021 AAGATGATTTGAACATATAGAGAGATTATGACAGCTGCAATTTGGCCGGAAGAGAGCCA 1080

QY 1081 CTGGTGCCTCAACCAACATCTATGCGGTTGAGGAGATGGGACCAAAATCTGGGTTTG 1140  
Db 1081 CTGGTGCCTCAACCAACATCTATGCGGTTGAGGAGATGGGACCAAAATCTGGGTTTG 1140

QY 1141 AAAAGTCAAGAGAGCTGGGAGAAAATACAGTATCTTAAATTAATGGAAGAGCTGTACACA 1200  
Db 1141 AAAAGTCAAGAGAGCTGGGAGAAAATACAGTATCTTAAATTAATGGAAGAGCTGTACACA 1200

QY 1201 TCCATTAACCTTGGGAAAATCTAAGAAAGCTGAAGCAAAATGTTAAAGAAATGACACA 1260  
Db 1201 TCCATTAACCTTGGGAAAATCTAAGAAAGCTGAAGCAAAATGTTAAAGAAATGACACA 1260

QY 1261 AACTGACAACTACAGAGAAAAGATCAGAGACAAAACGCTGGCTGAAAAATGCTTCTC 1320  
Db 1261 AACTGACAACTACAGAGAAAAGATCAGAGACAAAACGCTGGCTGAAAAATGCTTCTC 1320

QY 1321 CAGAGATGTGGAATATTATTAATCTGCAAGCAGAGAGCTTACAGATGATCTGCACAAACAT 1380  
Db 1321 CAGAGATGTGGAATATTATTAATCTGCAAGCAGAGAGCTTACAGATGATCTGCACAAACAT 1380

QY 1381 ATCAAAATAGTGAAGAAATATTTGCTCATTCATCAAAAGTCAAGAGCTGCTGTTATCCGG 1440  
Db 1381 ATCAAAATAGTGAAGAAATATTTGCTCATTCATCAAAAGTCAAGAGCTGCTGTTATCCGG 1440

QY 1441 ACTACTATTTGCAAAATGCGAGGCTGCTTACTAGAAATGAGTGAAGTGAAGTGAAGTGA 1500  
Db 1441 ACTACTATTTGCAAAATGCGAGGCTGCTTACTAGAAATGAGTGAAGTGAAGTGAAGTGA 1500

QY 1501 TCATTGCAAAAATGTTTCAGGACGCAATGTATGATATTTTACGAGAAATCATTCACAGA 1560  
Db 1501 TCATTGCAAAAATGTTTCAGGACGCAATGTATGATATTTTACGAGAAATCATTCACAGA 1560

QY 1561 CTACTCCCTTTAAAGATTCGCAAGGTTCTTAAACAGCAAGCAAGTGTGTCAGTGAAGA 1620  
Db 1561 CTACTCCCTTTAAAGATTCGCAAGGTTCTTAAACAGCAAGCAAGTGTGTCAGTGAAGA 1620

QY 1621 AGCAACCATCTTACATTTGAGAGACATGAAGTCTGAGTTAAAGATTTATCAGTTAATG 1680  
Db 1621 AGCAACCATCTTACATTTGAGAGACATGAAGTCTGAGTTAAAGATTTATCAGTTAATG 1680

QY 1681 GATTGAATTTGGGCTGCTCATTCATGATGCAAAAGAAATGATGATCTTGCAGATGAAA 1740  
Db 1681 GATTGAATTTGGGCTGCTCATTCATGATGCAAAAGAAATGATGATCTTGCAGATGAAA 1740

QY 1741 TGGGCTGGGTAACAAATACAAATTTCTTTTGAACATACCTGTTTCATGAACATC 1800  
Db 1741 TGGGCTGGGTAACAAATACAAATTTCTTTTGAACATACCTGTTTCATGAACATC 1800

QY 1801 AACTGATGAGCCCTTTCTTCTGCGCGTGCACCTTCTTACCTTGGCAAGAG 1860  
Db 1801 AACTGATGAGCCCTTTCTTCTGCGCGTGCACCTTCTTACCTTGGCAAGAG 1860

QY 1861 AGATTCAAACTTGGGCTCTCAGATGAAATGCTGAGTTTCTTAAAGAGATTAAGTAACTA 1920  
Db 1861 AGATTCAAACTTGGGCTCTCAGATGAAATGCTGAGTTTCTTAAAGAGATTAAGTAACTA 1920

QY 1921 GAAATATGATTAAGACTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980  
Db 1921 GAAATATGATTAAGACTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980

QY 1981 TACTTTCAGCAGATATGAATTTTACTGAAGATTAAGTATCTTGGTGTCTCAAT 2040  
Db 1981 TACTTTCAGCAGATATGAATTTTACTGAAGATTAAGTATCTTGGTGTCTCAAT 2040

QY 2041 GGGCATTCATAGAGGTTGATGAAGCTCATGCTTAAAAAATGATGACTCTTCTGTACA 2100  
Db 2041 GGGCATTCATAGAGGTTGATGAAGCTCATGCTTAAAAAATGATGACTCTTCTGTACA 2100



```
OY 2101 GGACTTATATGACCTTAAAGTCACCAACCATGACCTTCTGATTTCTGGAAGCCCACTGCAAA 2160
|||||
DB 2101 GGACTTATATGACCTTAAAGTCACCAACCATGACCTTCTGATTTCTGGAAGCCCACTGCAAA 2160
|||||
OY 2161 ATTCCCAAGAAAGCTGGCTTGTGTTGATTTGATTCATGCAAGAAATTTCTGCT 2220
|||||
DB 2161 ATTCCCAAGAAAGCTGGCTTGTGTTGATTTGATTCATGCAAGAAATTTCTGCT 2220
|||||
OY 2221 GGAAGATTTTGAAGAGGACATGGCAAGAGAGATATGGTTATGCAAGCTTTCACA 2280
|||||
DB 2221 GGAAGATTTTGAAGAGGACATGGCAAGAGAGATATGGTTATGCAAGCTTTCACA 2280
|||||
OY 2281 AAGACTTGAACCATTTTACTAAGAAAGCTTAAAAAGATGTAGAAAAGTCTTTACCTG 2340
|||||
DB 2281 AAGACTTGAACCATTTTACTAAGAAAGCTTAAAAAGATGTAGAAAAGTCTTTACCTG 2340
|||||
OY 2341 CTAAAGTTGAACAAATCTGAGATGGAATGAAATGATTCAGAGCAATATTTACAAGT 2400
|||||
DB 2341 CTAAAGTTGAACAAATCTGAGATGGAATGAAATGATTCAGAGCAATATTTACAAGT 2400
|||||
OY 2401 GGATTTTAAACAGGAATTTAAAGCCCTCAGTAAAGCTTCAAAAAGCAGTACCTCAGGCT 2460
|||||
DB 2401 GGATTTTAAACAGGAATTTAAAGCCCTCAGTAAAGCTTCAAAAAGCAGTACCTCAGGCT 2460
|||||
OY 2461 TTCTGAACATTTATGATGAGCACTTAAAGATGTTGTATACCATTTGCTACCTCATTAAGCCAC 2520
|||||
DB 2461 TTCTGAACATTTATGATGAGCACTTAAAGATGTTGTATACCATTTGCTACCTCATTAAGCCAC 2520
|||||
OY 2521 CAGATGAATGAATCTTATTAATTAACAGAGGCTTACAGCATTTGATACGTACAGAGC 2580
|||||
DB 2521 CAGATGAATGAATCTTATTAATTAACAGAGGCTTACAGCATTTGATACGTACAGAGC 2580
|||||
OY 2581 GGAACCTATCTCTTGTGACAAGCTACTGATTCCTGCGAAGACGGCAACAGAGTTC 2640
|||||
DB 2581 GGAACCTATCTCTTGTGACAAGCTACTGATTCCTGCGAAGACGGCAACAGAGTTC 2640
|||||
OY 2641 TGATTTCTCAGATGAGAGATGCTGACATCCATAGCAAAATCTGAAGATGCTCC 2700
|||||
DB 2641 TGATTTCTCAGATGAGAGATGCTGACATCCATAGCAAAATCTGAAGATGCTCC 2700
|||||
OY 2701 AGTTTCCCTCCAGAGACTTGTATGATCAATAAAGGGGAATTTGAGAGCAACGACTGG 2760
|||||
DB 2701 AGTTTCCCTCCAGAGACTTGTATGATCAATAAAGGGGAATTTGAGAGCAACGACTGG 2760
|||||
OY 2761 ATCATTTCAATGCAAGAGATCAGAGATTTCTGTTTTTACTGCTACAGAGCTGGAG 2820
|||||
DB 2761 ATCATTTCAATGCAAGAGATCAGAGATTTCTGTTTTTACTGCTACAGAGCTGGAG 2820
|||||
OY 2821 GATTAGTAATTAACCTTGGCATCTGTACACTGTATGATTTTGAATGCTGGAATC 2880
|||||
DB 2821 GATTAGTAATTAACCTTGGCATCTGTACACTGTATGATTTTGAATGCTGGAATC 2880
|||||
OY 2881 CACGAATGAATGCGAGGACAGCGGAGAGCTCATTAAGATTTGGACAGAAACAGGTTA 2940
|||||
DB 2881 CACGAATGAATGCGAGGACAGCGGAGAGCTCATTAAGATTTGGACAGAAACAGGTTA 2940
|||||
OY 2941 ATTTTATTCGGCTTACACAAAAGGATCAGTAGAAGATATTTCTTGAAGAGCCAAAGA 3000
|||||
DB 2941 ATTTTATTCGGCTTACACAAAAGGATCAGTAGAAGATATTTCTTGAAGAGCCAAAGA 3000
|||||
OY 3001 AGAAGATGCTGAGACATTTAGTAATTCAGAGAAATGACACAGAGAAAACCTGTTCC 3060
|||||
DB 3001 AGAAGATGCTGAGACATTTAGTAATTCAGAGAAATGACACAGAGAAAACCTGTTCC 3060
|||||
OY 3061 TGCATACAGGTTCAACCTCATCAACCTCTACACCTTTTAAATTAAGAAAGTATCAGCTA 3120
|||||
DB 3061 TGCATACAGGTTCAACCTCATCAACCTCTACACCTTTTAAATTAAGAAAGTATCAGCTA 3120
|||||
OY 3121 TTTTAAAGTTTGGTGTGAGAACTTTTAAAGAACCTGAAAGAGAGAAACAGAGCCCC 3180
|||||
DB 3121 TTTTAAAGTTTGGTGTGAGAACTTTTAAAGAACCTGAAAGAGAGAAACAGAGCCCC 3180
|||||
OY 3181 AGGAATGATATATGATGAATCTTGAAGAGAGCTGAAACCTGGGAAAATGAGCCAGGTC 3240
|||||
DB 3181 AGGAATGATATATGATGAATCTTGAAGAGAGCTGAAACCTGGGAAAATGAGCCAGGTC 3240
|||||
OY 3241 CATTCAGTGTGGGAGATGAGTGTCTTTCACAGTTCAAGGTGGCAACTTTTCCATATGG 3300
|||||
DB 3241 CATTCAGTGTGGGAGATGAGTGTCTTTCACAGTTCAAGGTGGCAACTTTTCCATATGG 3300
|||||
OY 3301 ATGAGATGATATTTGATTTGGAACCAAGAAATTTCAAGAAATTTGGCAAGAAATCATCC 3360
|||||
DB 3301 ATGAGATGATATTTGATTTGGAACCAAGAAATTTCAAGAAATTTGGCAAGAAATCATCC 3360
|||||
OY 3361 CAGATCCCAACGAGAGAGATAGAGAGAGGAGAAAGACAAAGAACTTGAAGAAATAT 3420
|||||
DB 3361 CAGATCCCAACGAGAGAGATAGAGAGAGGAGAAAGACAAAGAACTTGAAGAAATAT 3420
|||||
OY 3421 ACATGCTCCGAGATGAGAAACTGTGCAAAACAGATCAGCTTTAATGGAGTGAAGGAA 3480
|||||
DB 3421 ACATGCTCCGAGATGAGAAACTGTGCAAAACAGATCAGCTTTAATGGAGTGAAGGAA 3480
|||||
OY 3481 GACGAGTAGAGAGAGAGATATTTCTGATCTGATAGTACTCCATCAGAAAGAAAC 3540
|||||
DB 3481 GACGAGTAGAGAGAGAGATATTTCTGATCTGATAGTACTCCATCAGAAAGAAAC 3540
|||||
OY 3541 GGCCAAAAGAGCTGGAAGACCTCGAACCATTCCTCGAAGAAATATTAAGATTTAGTG 3600
|||||
DB 3541 GGCCAAAAGAGCTGGAAGACCTCGAACCATTCCTCGAAGAAATATTAAGATTTAGTG 3600
|||||
OY 3601 ATGCAAGATGAGGCGTTTATCAAGAGTTTCAAGAAATTTGGGCGCTGGAAGAGT 3660
|||||
DB 3601 ATGCAAGATGAGGCGTTTATCAAGAGTTTCAAGAAATTTGGGCGCTGGAAGAGT 3660
|||||
OY 3661 TAATGCTGTAGCTAGAGATGCTGAACTGGTGTATTAATCTGAGACAGACCTTAGACTT 3720
|||||
DB 3661 TAATGCTGTAGCTAGAGATGCTGAACTGGTGTATTAATCTGAGACAGACCTTAGACTT 3720
|||||
OY 3721 TGGGTGAACCTGTACATATATGATGATCATTAAGGCTTTAAGGACATTCATCTGGACAG 3780
|||||
DB 3721 TGGGTGAACCTGTGTACATATATGATGATCATTAAGGCTTTAAGGACATTCATCTGGACAG 3780
|||||
OY 3781 AAAGACAGAGAGTGAACCTTGGGAAAGTAAAGGCCCAAGCTTTCGATCTCAGAGTGC 3840
|||||
DB 3781 AAAGACAGAGAGTGAACCTTGGGAAAGTAAAGGCCCAAGCTTTCGATCTCAGAGTGC 3840
|||||
OY 3841 AGGTGAATGCAAAACTAGTCATCTCTCAGAAAGAGAGCTGGCACCACTGCACAAATCCA 3900
|||||
DB 3841 AGGTGAATGCAAAACTAGTCATCTCTCAGAAAGAGAGCTGGCACCACTGCACAAATCCA 3900
|||||
OY 3901 TTCTTCAGATTCAGAAAGAAAGAAATGTGTCATCCATGCCACACCAAGGCTGTC 3960
|||||
DB 3901 TTCTTCAGATTCAGAAAGAAAGAAATGTGTCATCCATGCCACACCAAGGCTGTC 3960
|||||
OY 3961 ACTTCGATATGATTTGGGCTTAAAGAAAGATGTCCTCAATCTGTTAGTGGATCTATGANT 4020
|||||
DB 3961 ACTTCGATATGATTTGGGCTTAAAGAAAGATGTCCTCAATCTGTTAGTGGATCTATGANT 4020
|||||
OY 4021 ATGGCTATGCGAGCTGGGAAATGATTAAGATGATCCAGATCTGCTTAACAGAGAGA 4080
|||||
DB 4021 ATGGCTATGCGAGCTGGGAAATGATTAAGATGATCCAGATCTGCTTAACAGAGAGA 4080
|||||
OY 4081 TTTTACCTGATGATCAGACAGAAACCCAGGCAAGGAGCTACAGACCTGCGAGACT 4140
|||||
DB 4081 TTTTACCTGATGATCAGACAGAAACCCAGGCAAGGAGCTACAGACCTGCGAGACT 4140
|||||
OY 4141 ACCTCATTTAAATTTACTGAATTAAGACCTTTGCAAGAAAGAGCAACAAAGGCTTGTGCTG 4200
|||||
DB 4141 ACCTCATTTAAATTTACTGAATTAAGACCTTTGCAAGAAAGAGCAACAAAGGCTTGTGCTG 4200
|||||
OY 4201 CAGGCAATTCAGAGAGAGAGAGACAGAAATTAAGAAAGATTAAGATGAGCTTCAAAA 4260
|||||
DB 4201 CAGGCAATTCAGAGAGAGAGAGACAGAAATTAAGAAAGATTAAGATGAGCTTCAAAA 4260
|||||
OY 4261 TAAAGAAAGAAATTAAGATGATCTTCCACCAACCTCAGAAAATCTGATGAAGATG 4320
|||||
```

Db 4261 TAAAAAGAAATTAAGAGTGTCTTCACCCACAACCTTCAGAAAAATCTGATGAGATG 4320  
QY 4321 ATGAGGAGGAGATPAACAAGSTAAATGAATCTGAAATTAAGAAAAATCTATAA 4380  
Db 4321 ATGAGGAGGAGATPAACAAGSTAAATGAATCTGAAATTAAGAAAAATCTATAA 4380  
QY 4381 AAATTCATGCTGGATACTCCAGTTCAATTAATTAATCTGCAACCGATGAACCACTTCTATCT 4440  
Db 4381 AAATTCATGCTGGATACTCCAGTTCAATTAATTAATCTGCAACCGATGAACCACTTCTATCT 4440  
QY 4441 CAGAGAAATCTGAAGAAGTCCATCAGAAAGACATTTAATGCTGCAAGAAAGATGAGGC 4500  
Db 4441 CAGAGAAATCTGAAGAAGTCCATCAGAAAGACATTTAATGCTGCAAGAAAGATGAGGC 4500  
QY 4501 CTGCAAGAGCAGCACTGAAAACAGCTGGATAGACAGAGAAAGGCGCTTCTGAAAGGAGC 4560  
Db 4501 CTGCAAGAGCAGCACTGAAAACAGCTGGATAGACAGAGAAAGGCGCTTCTGAAAGGAGC 4560  
QY 4561 AGCTGGAACATCTAGGAGAGTGTCTAATCAAAATTTGGGATCAGATTACAGAAATGCGTGA 4620  
Db 4561 AGCTGGAACATCTAGGAGAGTGTCTAATCAAAATTTGGGATCAGATTACAGAAATGCGTGA 4620  
QY 4621 AGGAGTACACAAATCCGAGCAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
Db 4621 AGGAGTACACAAATCCGAGCAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4680  
QY 4681 CCAAGTTTACAGATTTGATGCTGCAAGAAAGTGCACAAACCTTACAAACATCAATCAAA 4740  
Db 4681 CCAAGTTTACAGATTTGATGCTGCAAGAAAGTGCACAAACCTTACAAACATCAATCAAA 4740  
QY 4741 AGGCGCAAGAGTCTAGCAACACAAATGACCAAAACATTAGCAGCAATGTGAATACATG 4800  
Db 4741 AGGCGCAAGAGTCTAGCAACACAAATGACCAAAACATTAGCAGCAATGTGAATACATG 4800  
QY 4801 TAAATCAGAAATCCAGATGTGGAAGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
Db 4801 TAAATCAGAAATCCAGATGTGGAAGAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4860  
QY 4861 GGGAGAGTATTTCTGTATGAGATTTATCACAATACATGATCATCACAAGACAGAGC 4920  
Db 4861 GGGAGAGTATTTCTGTATGAGATTTATCACAATACATGATCATCACAAGAGGC 4920  
QY 4921 ATCAGGAGATGCTTACAGAAAGAGTACTCCAGGAAAGGCCATTTACGCTTCAGTA 4980  
Db 4921 ATCAGGAGATGCTTACAGAAAGAGTACTCCAGGAAAGGCCATTTACGCTTCAGTA 4980  
QY 4981 ATGGAAGAGATCAGAGAGAGTGGATCACTCAAAACAGAGACAGAGATCTACAGTGTATA 5040  
Db 4981 ATGGAAGAGATCAGAGAGAGTGGATCACTCAAAACAGAGACAGAGATCTACAGTGTATA 5040  
QY 5041 GTAAACATAGAAATGATGATGACCAAGAGACAGAGACAGAGACAGAGTCAAACTGGAAGAA 5100  
Db 5041 GTAAACATAGAAATGATGATGATGACCAAGAGACAGAGACAGAGTCAAACTGGAAGAA 5100  
QY 5101 ACTTAAAGACAGCGGGGTATTCAGATCAACCGCTCCATTCAGACACAGATACACT 5160  
Db 5101 ACTTAAAGACAGCGGGGTATTCAGATCAACCGCTCCATTCAGACACAGATACACT 5160  
QY 5161 CAGATCAACCGTTCACATTCAGAAATACAGCCATCAATTAATCTTCAGAGATTAATGATACC 5220  
Db 5161 CAGATCAACCGTTCACATTCAGAAATACAGCCATCAATTAATCTTCAGAGATTAATGATACC 5220  
QY 5221 ACTCAGAGTGGCAATGGACCAAGAGCTTCTGGTAGTGGCCGAGAGTCAACACTAGATC 5280  
Db 5221 ACTCAGAGTGGCAATGGACCAAGAGCTTCTGGTAGTGGCCGAGAGTCAACACTAGATC 5280  
QY 5281 AGAGGTCTCTTATGTTCAAGATCTCCCTAGAGACAGATCTTCATTTTAAACACTCAT 5340  
Db 5281 AGAGGTCTCTTATGTTCAAGATCTCCCTAGAGACAGATCTTCATTTTAAACACTCAT 5340  
QY 5341 CAGATCAACAAAGTACACCTGATACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5400  
Db 5341 CAGATCAACAAAGTACACCTGATACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5400

QY 5401 CATTTTCGACCTTCTTTTACCATATACAGTAAACTAACAGATTAATGCTTACAT 5460  
Db 5401 CATTTTCGACCTTCTTTTACCATATACAGTAAACTAACAGATTAATGCTTACAT 5460  
QY 5461 GACTTGAAGATATGAGACTGATATTTATCAGTAGCAGATATTTGATCTTCCAGGA 5520  
Db 5461 GACTTGAAGATATGAGACTGATATTTATCAGTAGCAGATATTTGATCTTCCAGGA 5520  
QY 5521 TGCAAGTCTATTTTCCCAAGAGAAAGAAATATTTTGTATTTTAAAGTTTATGCTGCAC 5580  
Db 5521 TGCAAGTCTATTTTCCCAAGAGAGAAAGAAATATTTTGTATTTTAAAGTTTATGCTGCAC 5580  
QY 5581 TGCTGCAAAATGTTGGACCTTTTAAAGAAATGGAAGATGTTTACTTTTAAAG 5640  
Db 5581 TGCTGCAAAATGTTGGACCTTTTAAAGAAATGGAAGATGTTTACTTTTAAAG 5640  
QY 5641 GACCTCAACACTGCCCTTTGAGACTGATCTTAAATCAAACTTCATGCAAGTGG 5700-  
Db 5641 GACCTCAACACTGCCCTTTGAGACTGATCTTAAATCAAACTTCATGCAAGTGG 5700  
QY 5701 TTTAGAGTGAACAGATTAATTAATGTTGTAATGAACCTTAAACACTGACCTGTG 5760  
Db 5701 TTTAGAGTGAACAGATTAATTAATGTTGTAATGAACCTTAAACACTGACCTGTG 5760  
QY 5761 CTTATGTTTCAAGAAAGATGGGGATTTATTTGTTTATTTCTGTTAGAGAACTCTC 5820  
Db 5761 CTTATGTTTCAAGAAAGATGGGGATTTATTTGTTTATTTCTGTTAGAGAACTCTC 5820  
QY 5821 AAGACCTTTGCTCACTTTCCAAAGCTTCTGTTTACATTTGACCTGCGACCTTGGC 5880  
Db 5821 AAGACCTTTGCTCACTTTCCAAAGCTTCTGTTTACATTTGACCTGCGACCTTGGC 5880  
QY 5881 GCTTTTCATCAAGCTTGAATTTTAAATCTGACTACAGTGTGTAATTAAGCCAGGA 5940  
Db 5881 GCTTTTCATCAAGCTTGAATTTTAAATCTGACTACAGTGTGTAATTAAGCCAGGA 5940  
QY 5941 TTTCTCCTGTTTGTATGATCAGTTAATGCTTTTATGAAACAAACAAACAAACAA 6000  
Db 5941 TTTCTCCTGTTTGTATGATCAGTTAATGCTTTTATGAAACAAACAAACAAACAA 6000  
QY 6001 CAATTTAAAAAACAACACAAACCAACAAATGCTGTAAATTTATGTAATTAAT 6060  
Db 6001 CAATTTAAAAAACAACACAAACCAACAAATGCTGTAAATTTATGTAATTAAT 6060  
QY 6061 AAATGAGCTTTTCCGACAGCTTTTGGCTGCTTCCCTCCCAACACTGAGGCT 6120  
Db 6061 AAATGAGCTTTTCCGACAGCTTTTGGCTGCTTCCCTCCCAACACTGAGGCT 6120  
QY 6121 TCTTTTCAAAAGTCAATTAATCAATGTTTAAATTAATCTCGATGAACTAGAT 6180  
Db 6121 TCTTTTCAAAAGTCAATTAATCAATGTTTAAATTAATCTCGATGAACTAGAT 6180  
QY 6181 GTAAAAAAGGGAAGGAATTTTATTCATTCATTAAGAGTCCCTTTTATGATGATCTT 6240  
Db 6181 GTAAAAAAGGGAAGGAATTTTATTCATTCATTAAGAGTCCCTTTTATGATGATCTT 6240  
QY 6241 TACATACCTGTTTGGTGTATTTATTTTATTTTCTATTTAACTGTCAGTGTG 6300  
Db 6241 TACATACCTGTTTGGTGTATTTATTTTATTTTCTATTTAACTGTCAGTGTG 6300  
QY 6301 TGATTTGTTTAATGAACAGTGAATATCCACTCTAAAGTGCCTCCGGAACCTTTTC 6360  
Db 6301 TGATTTGTTTAATGAACAGTGAATATCCACTCTAAAGTGCCTCCGGAACCTTTTC 6360  
QY 6361 AGGTGCAATGTTTAAAGGAAGAGTGTCTAAGGGAACCTTCAAAACCCAGATCA 6420  
Db 6361 AGGTGCAATGTTTAAAGGAAGAGTGTCTAAGGGAACCTTCAAAACCCAGATCA 6420  
QY 6421 GCCAAGATTCATTTAAATCAATTTGTTTCCCTTTAACTGGCAGTAATGATCAAT 6480  
Db 6421 GCCAAGATTCATTTAAATCAATTTGTTTCCCTTTAACTGGCAGTAATGATCAAT 6480

[illegible]

QY	661	AAACTGCCAAATGTCAGGGCTCAGGTCAGTACAGGAACGTGTTCTGATCTTGAATCGG	720
QY	661 <th>AAACATGCCAAATGTCAGGGCTCAGGCTCAGTACAGGAACGTGTTCTGATCTTGAATCGG</th> <th>720</th>	AAACATGCCAAATGTCAGGGCTCAGGCTCAGTACAGGAACGTGTTCTGATCTTGAATCGG	720
Db	661 <td>AAACATGCCAAATGTCAGGGCTCAGGCTCAGTACAGGAACGTGTTCTGATCTTGAATCGG</td> <td>720</td>	AAACATGCCAAATGTCAGGGCTCAGGCTCAGTACAGGAACGTGTTCTGATCTTGAATCGG	720
QY	721 <td>CGGAGATGGGGATTAAGACGTTGTGTAAGAAATGATCTCATCTATGACCCAAAAACA</td> <td>780</td>	CGGAGATGGGGATTAAGACGTTGTGTAAGAAATGATCTCATCTATGACCCAAAAACA	780
Db	721 <td>CGGAGATGGGGATTAAGACGTTGTGTAAGAAATGATCTCATCTATGACCCAAAAACA</td> <td>780</td>	CGGAGATGGGGATTAAGACGTTGTGTAAGAAATGATCTCATCTATGACCCAAAAACA	780

QY	781	AAGTCAAAAGCCGTTAAACCTCCAGCAGAAATTAAGCCAAAAAGTGGGAAAAAGACACAG	840
Db	781	AAGTCAAAAGCCGTTAAACCTCCAGCAGAAATTAAGCCAAAAAGTGGGAAAAAGACACAG	840
QY	841	GACGAAAGAAAGGCAACTTGATTCATCCAGAGAGAGAGAGAGACATGATGAAGTTATG	900
Db	841	GACGAAAGAAAGGCAACTTGATTCATCCAGAGAGAGAGAGAGACATGATGAAGTTATG	900
QY	901	ATAAGAGAGATCTCGTCGCGCAGGCAACAGTGAATTTAGTTACAAAGAGCTGAGAAA	960
Db	901	ATAAGAGAGATCTCGTCGCGCAGGCAACAGTGAATTTAGTTACAAAGAGCTGAGAAA	960
QY	961	CCAAAGACAGATTCGTGATGATTTCTGCGAAATTGTGCGAGAGATGCCACAGACTGAAG	1020
Db	961	CCAAAGACAGATTCGTGATGATTTCTGCGAAATTGTGCGAGAGATGCCACAGACTGAAG	1020
QY	1021	AAAGTGAATTTGAAACATATGAGAAAGTTATGACACGTCGATTTGGCCGAAAAGAGACCA	1080
Db	1021	AAAGTGAATTTGAAACATATGAGAAAGTTATGACACGTCGATTTGGCCGAAAAGAGACCA	1080
QY	1081	CTGGTGCCCTCAACCCACCATCTATCCGTTGAGGACAGATGTCACCCAAATGCTGGGTTG	1140
Db	1081	CTGGTGCCCTCAACCCACCATCTATCCGTTGAGGACAGATGTCACCCAAATGCTGGGTTG	1140
QY	1141	AAAAGTCAAAAGGAGCTGGGGAATTAACATGATCTTATTAATGGAAGAGCTGTCACACA	1200
Db	1141	AAAAGTCAAAAGGAGCTGGGGAATTAACATGATCTTATTAATGGAAGAGCTGTCACACA	1200
QY	1201	TCCATTAACACTTGGGAAACCTGAAGAAACGCTGACGACAAACAAATTTAAAGGAATGAACA	1260
Db	1201	TCCATTAACACTTGGGAAACCTGAAGAAACGCTGACGACAAACAAATTTAAAGGAATGAACA	1260
QY	1261	AACGTGACAACTACAGAAAAAGGATCAGAGACAAACCGCTGCGTGAATAATGCTTCTC	1320
Db	1261	AACGTGACAACTACAGAAAAAGGATCAGAGACAAACCGCTGCGTGAATAATGCTTCTC	1320
QY	1321	CAGAGAGATGGGAATTTATTAACCTGCGCAGAGAGCTTCAATCATCTGCAACAAACAT	1380
Db	1321	CAGAGAGATGGGAATTTATTAACCTGCGCAGAGAGCTTCAATCATCTGCAACAAACAT	1380
QY	1381	ATCAAAATAGTGGAAAGAAATATTGCTCATTTCAAAATCAAAAGCTCAGACGCTGATATCCGG	1440
Db	1381	ATCAAAATAGTGGAAAGAAATATTGCTCATTTCAAAATCAAAAGCTCAGACGCTGATATCCGG	1440
QY	1441	ACTACTATGTGAAATGGCAGGGTCTGCTTACTCAGAAATGTAAGCTGGGAAAGATGCTGTC	1500
Db	1441	ACTACTATGTGAAATGGCAGGGTCTGCTTACTCAGAAATGTAAGCTGGGAAAGATGCTGTC	1500
QY	1501	TCATTTGCCAAAAAGTTTCAGGCAACGATGATGATTAATTTAGCAAGAAATCATCTCAAGA	1560
Db	1501	TCATTTGCCAAAAAGTTTCAGGCAACGATGATGATTAATTTAGCAAGAAATCATCTCAAGA	1560
QY	1561	CTACTCCCTTTAAGGACTGCAAGGTTCTAAACACAGAACCAAGATTTGTTGCATGAAGA	1620
Db	1561	CTACTCCCTTTAAGGACTGCAAGGTTCTAAACACAGAACCAAGATTTGTTGCATGAAGA	1620
QY	1621	AGCAACCATCTTACATTTGGAGAGACATGAAGAGTCTGGAAGTTAAGACATTAATCAGTTAAATG	1680
Db	1621	AGCAACCATCTTACATTTGGAGAGACATGAAGAGTCTGGAAGTTAAGACATTAATCAGTTAAATG	1680
QY	1681	GATTGAATTTGGCTGCGCTCATTTCAATGCTGCAAGGAAATGTTGTATTTCTTGACAGTGAATC	1740
Db	1681	GATTGAATTTGGCTGCGCTCATTTCAATGCTGCAAGGAAATGTTGTATTTCTTGACAGTGAATC	1740
QY	1741	TGGGCTGGGTTAAACAAATACAAACAATTTCTTTCTGAACTACCTGTTTCATGAACATC	1800
Db	1741	TGGGCTGGGTTAAACAAATACAAACAATTTCTTTCTGAACTACCTGTTTCATGAACATC	1800
QY	1801	AACGTATAGGCCCTTTTCTTCTGCGGCTGCGACATTTCTACCTTGACATCTTGGCAAGAG	1860
Db	1801	AACGTATAGGCCCTTTTCTTCTGCGGCTGCGACATTTCTACCTTGACATCTTGGCAAGAG	1860

QY	1861	AGATTCAAACTTGGGCTCTCCACATGATGATGCTGTACTTACTTACGAGATATTAACCTACTA	1920
Db	1861	AGATTCAAACTTGGGCTCTCCACATGATGATGCTGTACTTACTTACGAGATATTAACCTACTA	1920
QY	1921	GAATATGATATAGACACTCATGATGATGCATCCACAGACTAAAGCATTAAGTTTAACA	1980
Db	1921	GAATATGATATAGACACTCATGATGATGCATCCACAGACTAAAGCATTAAGTTTAACA	1980
QY	1981	TACTTTCGACGACATATGAATTTTACTGAGAGATTAAGTCATTCCTGTGTGGTCTCAATT	2040
Db	1981	TACTTTCGACGACATATGAATTTTACTGAGAGATTAAGTCATTCCTGTGTGGTCTCAATT	2040
QY	2041	GGGCACTTCATAGGAGTTGATGAAGCTCATCGTTTAAAAAATATACCTCTCTCGTACA	2100
Db	2041	GGGCACTTCATAGGAGTTGATGAAGCTCATCGTTTAAAAAATATACCTCTCTCTGTRCA	2100
QY	2101	GGACTTTAATAGACTTTTAAGTCCAACCATCGACTTCGTATTACTGGAAGCCCACTGCAAA	2160
Db	2101	GGACTTTAATAGACTTTTAAGTCCAACCATCGACTTCGTATTACTGGAAGCCCACTGCAAA	2160
QY	2161	ATTCCTTCACAAAGACTGTGGTCTTTGTGTGATTTTCACTATGCGCAAAAAATTTTCCTCT	2220
Db	2161	ATTCCTTCACAAAGACTGTGGTCTTTGTGTGATTTTCACTATGCGCAAAAAATTTTCCTCT	2220
QY	2221	GGGAAGAATTTTGAAGAGAGACATGGCAAAAGAGAGATGTGTATCCAAAGTCCTTGACA	2280
Db	2221	GGGAAGAATTTTGAAGAGAGACATGGCAAAAGAGAGATGTGTATCCAAAGTCCTTGACA	2280
QY	2281	AAGAGCTTGAACCATTTTCTACTAAGAAAGTTTAAAAAGATATACAAAAGCTTTACTCTG	2340
Db	2281	AAGAGCTTGAACCATTTTCTACTAAGAAAGTTTAAAAAGATATACAAAAGCTTTACTCTG	2340
QY	2341	CTAAGTGTGAACAATTCCTAGAGATGGAATGGAATGAGTGCATTCGAGAGCAATATTACAAGT	2400
Db	2341	CTAAGTGTGAACAATTCCTAGAGATGGAATGGAATGAGTGCATTCGAGAGCAATATTACAAGT	2400
QY	2401	GGATTTTACCAAGSATTATTAAGCCCTCACTAAGGCTTCAAAAGGCACTACCTCAGGCT	2460
Db	2401	GGATTTTACCAAGSATTATTAAGCCCTCACTAAGGCTTCAAAAGGCAAGCTACCTCAGGCT	2460
QY	2461	TTTCGACACTATATGATGAGACTTAAGAAAGTGTGTACCATGTGCATCTCATTTAAAGCAC	2520
Db	2461	TTTCGACACTATATGATGAGACTTAAGAAAGTGTGTACCATGTGCATCTCATTTAAAGCAC	2520
QY	2521	CAGATGATATGAATTCCTATATTAATAACAGAGAGCCTTACAGCATTTGATACGTAGCAGCG	2580
Db	2521	CAGATGATATGAATTCCTATATTAATAACAGAGAGCCTTACAGCATTTGATACGTAGCAGCG	2580
QY	2581	GGAAACTAATCCCTCTTGACAAAGCTCATGATTCGTGCGAGAAAGTGTGGCAACAGAGCTC	2640
Db	2581	GGAAACTAATCCCTCTTGACAAAGCTCATGATTCGTGCGAGAAAGTGTGGCAACAGAGCTC	2640
QY	2641	TGATTTTCTCTCAGATGAGTGTAGAGATCTGTGACATCTCTAGCAGAAATTCGAAATATGCC	2700
Db	2641	TGATTTTCTCTCAGATGAGTGTAGAGATCTGTGACATCTCTAGCAGAAATTCGAAATATGCC	2700
QY	2701	AGTTTCCCTTCAGAGACTTGATGATGATCAATAAAGGGGAATTGGAGAGCAAGCATCTGG	2760
Db	2701	AGTTTCCCTTCAGAGACTTGATGATGATCAATAAAGGGGAATTGGAGAGCAAGCATCTGG	2760
QY	2761	ATCATTTCAATGAGAGATCAAGAGATTCCTGTTTTTACTGTCTCTCAAGAGCTGGAG	2820
Db	2761	ATCATTTCAATGAGAGATCAAGAGATTCCTGTTTTTACTGTCTCTCAAGAGCTGGAG	2820
QY	2821	GATTATGATTAACCTTGGCATCTGCTGACACTGTAGATTATTTTGTATTCGTACCTGGAATC	2880
Db	2821	GATTATGATTAACCTTGGCATCTGCTGACACTGTAGATTATTTTGTATTCGTACCTGGAATC	2880
QY	2881	CACAGATGATCTGACAGGACACAGGCGCAGAGCTCATAGAATTTGGACAGAAACACAGGTA	2940
Db	2881	CACAGATGATCTGACAGGACACAGGCGCAGAGCTCATAGAATTTGGACAGAAACACAGGTA	2940
QY	2941	ATATTATTCGGCTAGTACACAAAGAGATCAAGTATGAGAGACATATTTCTTAAGAGCCACAGA	3000

```

|||||
Db 2941 ATATTATCGGCTAGTCAAAAAGATAGATAGATATCTTGAAGCAAGA 3000
Oy 3001 ACAGAGATGCTGACACATTTAGTAATTCAGAGATGACACAGAGAAACCTGTC 3060
Db 3001 ACAGAGATGCTGACACATTTAGTAATTCAGAGATGACACAGAGAAACCTGTC 3060
Oy 3061 TGCATACAGGTTCACTCCATCAAGCTTACACCTTTTAAAGAAAGTTATACAGTA 3120
Db 3061 TGCATACAGGTTCACTCCATCAAGCTTACACCTTTTAAAGAAAGTTATACAGTA 3120
Oy 3121 TTTTGAAGTTGCTGTAGAGACTTTTAAAGACCTGAAGAGAGAGAGAGAGCC 3180
Db 3121 TTTTGAAGTTGCTGTAGAGACTTTTAAAGACCTGAAGAGAGAGAGAGAGCC 3180
Oy 3181 AGGAATGATATAGTGAATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Db 3181 AGGAATGATATAGTGAATCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
Oy 3241 CATTGACTGAGGAGATGAGATGCTTCAAGTTCAGAGTTCAGAGTTCAGATAGG 3300
Db 3241 CATTGACTGAGGAGATGAGATGCTTCAAGTTCAGAGTTCAGAGTTCAGATAGG 3300
Oy 3301 ATGAAGATGATATGATGAGTTCAGACCAAGAAATTCAGAAATTCAGAAATCACC 3360
Db 3301 ATGAAGATGATATGATGAGTTCAGACCAAGAAATTCAGAAATTCAGAAATCACC 3360
Oy 3361 CAGAAATCCCAAGAGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Db 3361 CAGAAATCCCAAGAGAGAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
Oy 3421 ACATGCTCCGAGAGATGAGAAATCTGCAGAAACAGATAGCTTTATAGGAGATGAGAA 3480
Db 3421 ACATGCTCCGAGAGATGAGAAATCTGCAGAAACAGATAGCTTTATAGGAGATGAGAA 3480
Oy 3481 GACGAGATGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3540
Db 3481 GACGAGATGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 3540
Oy 3541 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Db 3541 GGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
Oy 3601 ATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Db 3601 ATGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
Oy 3661 TTAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3720
Db 3661 TTAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3720
Oy 3721 TGGGTGAATCTTGTACATTAATGATGATGATGATGATGATGATGATGATGATG 3780
Db 3721 TGGGTGAATCTTGTACATTAATGATGATGATGATGATGATGATGATGATGATG 3780
Oy 3781 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
Db 3781 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3840
Oy 3841 AGGTGAATGCAAAAGTCAATCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Db 3841 AGGTGAATGCAAAAGTCAATCTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900
Oy 3901 TTTCTTCAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Db 3901 TTTCTTCAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3960
Oy 3961 ACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4020
Db 3961 ACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4020
Oy 4021 ATGGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Db 4021 ATGGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080

|||||
Db 4021 ATGGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4080
Oy 4081 TTTTACGTGATGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Db 4081 TTTTACGTGATGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4140
Oy 4141 ACCTGATTAATTCATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200
Db 4141 ACCTGATTAATTCATGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4200
Oy 4201 CAGGCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260
Db 4201 CAGGCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4260
Oy 4261 TAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320
Db 4261 TAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4320
Oy 4321 ATGAGAGAGAGATACAGAG 4341
Db 4321 ATGAGAGAGAGATACAGAG 4341

RESULT 3
D14316 2292 bp mRNA linear VRT 03-FEB-1999
LOCUS delta-crystallin enhancer binding protein, complete sequence.
DEFINITION D14316
ACCESSION D14316
VERSION D14316.1 GI:391639
KEYWORDS ORF2.
SOURCE Gallus gallus (library: lambda gtl1) 13 day embryo lens cDNA to mRNA, clone ffl1.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 2292)
AUTHORS Funahashi,J.
TITLES Direct Submission
JOURNAL Submitted (29-JAN-1993) Jun-ichi Funahashi, Inst. Devel., Aging and
Cancer, Tohoku Univ., 4-1 Seiryu-machi Aoba-ku, Sendai 980-77,
Japan, Tel:022-272-9499, Fax:022-272-3982)
2 (bases 1 to 2292)
AUTHORS Funahashi,J., Sekido,R., Murai,K., Kamachi,Y. and Kondoh,H.
TITLES Delta-crystallin enhancer binding protein delta EPI is a zinc
finger-homeodomain protein implicated in postgastrulation
embryogenesis
JOURNAL Development 119 (2), 433-446 (1993)
MEDLINE 9411644
REFERENCE 3 (bases 1 to 2292)
AUTHORS Funahashi,J.
JOURNAL Unpublished (1994)
FEATURES
source
location/Qualifiers
1..2292
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="lens"
/clone_lib="lambda gtl1"
/dev_stage="13 day embryo"
/note="Delta-gal fusion protein binds to blocks 10 to 3 of
HN fragment of delta-crystallin enhancer."
/codon_start=1
/protein_id="BA03262.1"
/db_xref="GI:391640"
/translaton="MELKKCNHLYLRPPDNEFYFNKOALDLOHLIRSSKLLILDKL
LIRLRBRNRYLIRFOSKVRMLDILAEYLYKROPFORLDGSKGLRKOALDHFNAEG
SEDFCLSTRAGGLGINTLASADIVYIEPSDNRPODLOAOBARHRIQKQKOVNRYL
VWGSVEEDILIRAKKMYLDHLYVIOBMDTTCVYLRGSSPSSSPFNKRELSALIK
FGAEELFKEPGEEROPQMDIDELIKRAEIRENRPGLTVGDELLSPFVNPFSNM
EDDILIEPNSRMKEIIPESORRIIEPRROKTELLEYMLPWRMCAKQISFNGSG
GRSRSRRTSGSDSDSTTERKPKRKGRPRITRENIKFSDAIRRFISIKKKGCP

```







DEFINITION Sequence 3 from Patent WO9639505.  
ACCESSION A58684  
VERSION A58684.1 GI:3714247  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 153)  
AUTHORS Griffiths,R. and Tilwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 3 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES Other publication AU 5906996 961224.  
location/Qualifiers  
source 1..153  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 58 a 40 c 31 g 24 t  
ORIGIN

Query Match 1.6%; Score 103; DB 6; Length 153;  
Best Local Similarity 100.0%; Pred. No. 3.3e-40;  
Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4100 CAAGAAACCCGACCAAGACAGCTACAGACCCGTCAGACTCTCATTAATTACTGAA 4159  
|||||  
Db 51 CAAGAAACCCGACCAAGACAGCTACAGACCCGTCAGACTCTCATTAATTACTGAA 110  
|||||

Qy 4160 TAAAGACCTTGCAAGAAAGAGCACAAGGCTTGCTGCTGCA 4202  
|||||  
Db 111 TAAAGACCTTGCAAGAAAGAGCACAAGGCTTGCTGCTGCA 153  
|||||

RESULT 8  
AF181824  
LOCUS AF181824 2754 bp mRNA linear VRT 07-AUG-2001  
DEFINITION Aegolius funereus chromosome w chromodomain helicase DNA binding  
ACCESSION AF181824  
VERSION AF181824  
KEYWORDS protein 1 (CHD1W) mRNA, partial cds.  
SOURCE AF181824.1 GI:5917748  
ORGANISM Aegolius funereus.  
REFERENCE Aegolius funereus  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Strigiformes; Strigidae; Aegolius.  
TITLE 1 (bases 1 to 2754)  
AUTHORS Fridolfsson,A.K. and Ellegren,H.  
JOURNAL Molecular evolution of the avian CHD1 genes on the Z and W sex  
MEDLINE chromosomes  
20384598 Genetics 155 (4), 1903-1912 (2000)  
2 (bases 1 to 2754)  
AUTHORS Fridolfsson,A.K. and Ellegren,H.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-1999) Department of Evolutionary Biology,  
Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D,  
Uppsala S-75236, Sweden  
location/Qualifiers  
source 1..2754  
/organism="Aegolius funereus"  
/db\_xref="taxon:103375"  
/chromosome="W"  
<1..>2754  
/gene="CHD1W"  
<1..>2754  
/gene="CHD1W"  
/codon\_start=1  
/product="chromodomain helicase DNA binding protein 1"  
/protein\_id="A056021.1"  
/db\_xref="GI:5917753"  
/translation="ELTDDLKQYQYIERIIANSNOKSAAGPYDYCKWGLPYSECS

WEDGALIAKKFQACIDIEYFSRNSKTTPEKCKIILKORPFAVLAKKOPSTYIGCHESELE  
LRDYOLNGLMIAHSMCKNSCILIADEMGDKTIQITISFLNPTFHHQILGPPLVLP  
LSTLTSWOREIOTMAPOMNSVVLGDIISNMARTEHMHDPQRLKFNLLTTEYL  
LKDSFLGLNWEIYGDVBAHRLKNDOSLLYKTLIDPKSVMHRLIGTPILONSLKEIM  
SLHFTIMPEKSSMEDEEERHGGREGVYSLKLEPELTLRRVKKDVEKSLPAKVEO  
ILNEMNALOKOYTKITLITNRYKALSGSGSGTSGLTINMELKCCNCTYLIKPPDD  
NEFYNKOEALQHLIRSSGKLLIDKLLIRERGNRYLIFSQVMRLDIIAEYLTAKO  
FPPRLDGSJIGELRKQALDHFNABEGSEDFCLSTFAGGLINLASADTVILFSDW  
NPONDLOAQARARIGOKOVNIYRLVTKGSVEDILERAKKMIVLDHVIQOMDTTG  
KYVHTGTSPTSSSTPEKKELSAILKGAELFKPEPEEPEEPEOIMDIEILKRAETR  
ENBERGLTVGDELLSOIKVANSFMDEDDIELPEONFKMEETITPEVORRIEERER  
OKELEELIYMPHRRNCAKOJSPFNGSBEKRSRKSYSDDSDSISEKKRPKRPRPTI  
PRENKEFSDAEIRRFKTSYKFKGPELRIADARQELDKSETUDRLGELVHYGC  
IKALKDMSFGQERAGRLKGVKPTPEFISGVONAKIYISHBELAPLHKSLSDPED  
RRRYVIFCHTKAAHFIDIMCKEDDSNLIIGIYBYGYGSMWIMIMDPLSLTKILPDD  
PDKK"

BASE COUNT 943 a 461 c 623 g 727 t  
ORIGIN

Query Match 1.3%; Score 83; DB 5; Length 2754;  
Best Local Similarity 100.0%; Pred. No. 3.5e-30;  
Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2919 ATTGGACAGAAACACGTTAATATTATTCGCTAGTCACAAAAGATCAGTGAAGAA 2978  
|||||  
Db 1567 ATTGGACAGAAACACGTTAATATTATTCGCTAGTCACAAAAGATCAGTGAAGAA 1626  
|||||

Qy 2979 GATATTCCTTGAAGAGCCAAAGAA 3001  
|||||  
Db 1627 GATATTCCTTGAAGAGCCAAAGAA 1649  
|||||

RESULT 9  
AF006513  
LOCUS AF006513 5947 bp mRNA linear PRI 27-NOV-1997  
DEFINITION Homo sapiens CHD1 mRNA, complete cds.  
ACCESSION AF006513  
VERSION AF006513.1 GI:2645428  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM human.  
REFERENCE Homo sapiens  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 5947)  
AUTHORS Woodage,T., Bastral,M.A., Baxevanis,A.D., Hietter,P. and Collins,F.S.  
JOURNAL Characterization of the CHD family of proteins  
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 94 (21), 11472-11477 (1997)  
97470991  
2 (bases 1 to 5947)  
AUTHORS Woodage,T.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUN-1997) Laboratory of Gene Transfer, National Human  
Genome Research Institute, National Institutes of Health, 49  
Convent Drive, Bethesda, MD 20892-4442, USA  
location/Qualifiers  
source 1..5947  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/map="5q15-21; near WT-5811"  
1..5947  
/gene="CHD1"  
164..5293  
/gene="CHD1"  
/codon\_start=1  
/product="CHD1"  
/protein\_id="AAB87381.1"  
/db\_xref="GI:2645429"  
/translation="MNCHSDSESVRNSSGSSQSDDDSGSAGSGSGSSGSSSSSSSS  
SOGSSPSDSSGSSGSSGSSSESDPTSRNKVOAKPPYVDGAEFKSSPSTIAYVRATL  
KKOQOQOQOQOHAASSNSGSEEDSSSDSDSSSVKRRKKKHDEDMOWSGSSSPSS  
GSDSESEEREKSSCDETSDYEPKKNKVRKPNOKNSKNGKTLIGOKRKQIDSSSE





QY 5836 TTTCGAAGCTACTGTTTACATGTGACACGACACCTTGGCGCTTTTCATCACAAG 5895  
|||||  
Db 133120 TTTCGAAGCTACTGTTTACATGTGACACGACACCTTGGCGCTTTTCATCACAAG 133179  
QY 5896 CTGGAATATTAAATTCGTGAC 5917  
|||||  
Db 133180 CTGGAATATTAAATTCGTGAC 133201

RESULT 12  
AC021449  
LOCUS  
DEFINITION  
AC021449 143079 bp DNA linear HTG 10-SEP-2000  
Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered  
pieces.  
AC021449 GI:10047806  
VERSION  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
ORGANISM  
human.  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS  
1 (bases 1 to 143079)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens, clone RP11-58M12  
JOURNAL  
Unpublished  
2 (bases 1 to 143079)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,  
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,  
Choezel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,  
Deaileano, K., Dewar, K., Domino, M., Doyle, M., Fenesfor, J.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,  
Landers, T., Leoczek, J., Levine, R., Lieu, C., Liu, G., Locke, K.,  
Macdonald, P., Margulis, N., McEwan, P., McGuck, A., McKenna, K.,  
McPheters, R., Melgrim, J., Meneus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
Pierre, N., Pisanl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Turrell, A., Vassiliev, H., Vtel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
Zimmer, A. and Zody, M.

TITLE  
JOURNAL  
COMMENT  
Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:7407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: L5154  
Center clone name: 58\_M\_12

Summary Statistics  
Sequencing vector: MJ3; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q30  
Consensus quality: 140814 bases at least Q20  
Insert size: 14400; agarose-fp  
Insert size: 142179; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 38820: contig of 38820 bp in length  
\* 38821 38920: gap of 100 bp  
\* 38921 40411: contig of 1491 bp in length  
\* 40412 40511: gap of 100 bp  
\* 40512 43279: contig of 2768 bp in length  
\* 43280 43379: gap of 100 bp  
\* 43380 46905: contig of 3526 bp in length  
\* 46906 47005: gap of 100 bp  
\* 47006 51830: contig of 4825 bp in length  
\* 51831 51930: gap of 100 bp  
\* 51931 62619: contig of 10669 bp in length  
\* 62620 62719: gap of 100 bp  
\* 62720 75408: contig of 12689 bp in length  
\* 75409 75508: gap of 100 bp  
\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length  
\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.

Location/Qualifiers  
1. 143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-58M12"  
/clone\_id="RP11-58M12 Human Male BAC"  
1. 38820  
/note="assembly-fragment  
clone\_end:SP6  
vector\_side:left"  
38921. 40411  
/note="assembly-fragment"  
40512. 43279  
/note="assembly-fragment"  
43380. 46905  
/note="assembly-fragment"  
47006. 51830  
/note="assembly-fragment"  
51931. 62619  
/note="assembly-fragment"  
62720. 75408  
/note="assembly-fragment"  
75509. 92516  
/note="assembly-fragment"  
92617. 106409  
/note="assembly-fragment"  
106510. 143079  
/note="assembly-fragment  
clone\_end:T7  
vector\_side:right"

BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

Query Match 1.2%: Score 82; DB 2; Length 143079;  
Best Local Similarity 100.0%; Pred. No. 7.7e-30;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCGAAGCTACTGTTTACATGTGACACGACACCTTGGCGCTTTTCATCACAAG 5895  
|||||  
Db 133109 TTTCGAAGCTACTGTTTACATGTGACACGACACCTTGGCGCTTTTCATCACAAG 133108  
QY 5896 CTGGAATATTAAATTCGTGAC 5917  
|||||  
Db 1331079 CTGGAATATTAAATTCGTGAC 1331100

RESULT 13  
AC008531

LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE.  
ACCESSION AC008531  
VERSION AC008531.3 GI:12830078  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 5  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 145659)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 14, 2001 this sequence version replaced gi:7528342.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 369535  
Center Clone Name: CIT-HSPC\_480B11  
-----  
Summary Statistics  
Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 148000; pulse field gel estimation  
Estimated insert size: 145059; sum-of-coverage estimation  
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation  
NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 56175 56174: contig of 56174 bp in length  
\* 56275 56274: gap of unknown length  
\* 100875 100874: contig of 44600 bp in length  
\* 100975 100974: gap of unknown length  
\* 113128 113127: contig of 12153 bp in length  
\* 113228 113227: gap of unknown length  
\* 118191 118190: contig of 4963 bp in length  
\* 118281 118280: gap of unknown length  
\* 119685 119684: contig of 1404 bp in length  
\* 119785 119784: gap of unknown length  
\* 123298 123297: contig of 3503 bp in length  
\* 123398 123397: gap of unknown length  
\* 123398 145659: contig of 22262 bp in length.  
Location/Qualifiers  
1. 145659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-480B11"  
/clone\_1lb="Caltech human BAC library C"  
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others  
ORIGIN

Query Match 1.2%; Score 82; DB 2; Length 145659;  
Best Local Similarity 100.0%; Pred. No. 7.7e-30;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5836 TTTCACAAAGCTACTGTTTACATGTGTACACCTGGACACCTTGGCGCTTTTCACACAG 5895  
|||||  
DB 61620 TTTCACAAAGCTACTGTTTACATGTGTACACCTGGACACCTTGGCGCTTTTCACACAG 61679  
OY 5896 CTTGAATATTTAAATTCGTAC 5917  
|||||  
DB 61680 CTTGAATATTTAAATTCGTAC 61701  
RESULT 14  
AC022121 219258 bp DNA linear PRI 30-AUG-2001  
LOCUS AC022121/c  
DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.  
ACCESSION AC022121  
VERSION AC022121.6 GI:15375145  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 219258)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 219258)  
REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
AUTHORS DOE Joint Genome Institute  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
On Aug 30, 2001 this sequence version replaced gi:15148108.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated total Number of Errors is 0.4.  
STS Content:  
WT-5811 G04974  
WT-13675 G23101  
SHGC-58345 G38487  
SHGC-103595 G57841.  
Location/Qualifiers  
1. 219258  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTD-2007H13"  
BASE COUNT 71954 a 42062 c 40933 g 64309 t  
ORIGIN

Query Match 1.2%; Score 82; DB 9; Length 219258;  
Best Local Similarity 100.0%; Pred. No. 7.4e-30;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5836 TTTCACAAAGCTACTGTTTACATGTGTACACCTGGACACCTTGGCGCTTTTCACACAG 5895  
|||||  
DB 74578 TTTCACAAAGCTACTGTTTACATGTGTACACCTGGACACCTTGGCGCTTTTCACACAG 74519  
OY 5896 CTTGAATATTTAAATTCGTAC 5917  
|||||  
DB 74518 CTTGAATATTTAAATTCGTAC 74497

## RESULT 15

AF181827

AF181827 2754 bp mRNA linear VRT 07-AUG-2001  
Nymphicus hollandicus chromosome W chromodomain helicase DNA

binding protein 1 (CHD1W) mRNA, partial cds.

AF181827.1 GI:5917751

## ACCESSION

AF181827

## VERSION

AF181827.1

## KEYWORDS

AF181827.1

## SOURCE

AF181827.1

## ORGANISM

AF181827.1

## REFERENCE

AF181827.1

## AUTHORS

AF181827.1

## TITLE

AF181827.1

## JOURNAL

AF181827.1

## MEDLINE

AF181827.1

## REFERENCE

AF181827.1

## AUTHORS

AF181827.1

## TITLE

AF181827.1

## JOURNAL

AF181827.1

## MEDLINE

AF181827.1

## REFERENCE

AF181827.1

## AUTHORS

AF181827.1

## TITLE

AF181827.1

## JOURNAL

AF181827.1

## MEDLINE

AF181827.1

## REFERENCE

AF181827.1

## AUTHORS

AF181827.1

## TITLE

AF181827.1

Search completed: August 3, 2002, 02:12:57  
Job time: 42547 sec

AF181827 2754 bp mRNA linear VRT 07-AUG-2001  
Nymphicus hollandicus chromosome W chromodomain helicase DNA  
binding protein 1 (CHD1W) mRNA, partial cds.

AF181827.1 GI:5917751

cockatiel.  
Nymphicus hollandicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Psittaciformes; Cacatidae;  
Nymphicus.

1 (bases 1 to 2754)  
Fridolfsson, A.K. and Ellegren, H.  
Molecular evolution of the avian CHD1 genes on the Z and W sex  
chromosomes  
Genetics 155 (4), 1903-1912 (2000)

2 (bases 1 to 2754)  
Fridolfsson, A.K. and Ellegren, H.  
Direct Submission  
Submitted (30-AUG-1999) Department of Evolutionary Biology,  
Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D,  
Uppsala S-75236, Sweden  
Location/Qualifiers  
1..2754  
/organism="Nymphicus hollandicus"  
/strain="Lutino cockatiel"  
/db\_xref="taxon:13180"  
/chromosome="W"  
<1..>2754  
/gene="CHD1W"  
<1..>2754  
/gene="CHD1W"  
/codon\_start=1  
/product="chromodomain helicase DNA binding protein 1"  
/db\_xref="GI:5917756"

/translation="ETDDHKOYQIVERRIANSNOKSEAGYPPDYCKKQGLPYSECS  
WEDGALIAKFOACIDEYFSRNLSTKTPEDCKILRQPRFVALKQPSYIGHSLE  
LRDQOLNGMLNLAHWCNCSILADEMGKTIQIISFNMYLFHEHOLGPELVLP  
LSTLSMREIQTMAPQMAVYIGDITSRNMRTHMMHPTKRLFNILLTYEIL  
LKDSFLGIMWFIQVDEAHRLKNDSDILYKLLIDKSNHRLITGTPLONSLREIM  
SILHFMPEKSSWDEDEEHGKRGREGYASLKELEPLLRVKKDYKSLPARYEQ  
ILRMESALQKQYKWLITRNLYALSNGSKSTSGFLNIMELKCCNHCYLITPDD  
NEFYRKQEAQLIRSSGKLLIDKLLIRLREGNRYLIFSQVRLMLDLAEYLKTRQ  
FPQRLDGSIKGELRKQALDFNABEGSEDFPLSTRAGLGINLASADTVVIFDSW  
NPQNDLQAQARAHIGOKKOVNRYRVTKGSEVEDILERAKKMVLADHLYIQMDTG  
KTVLHGTSPSSSTPNKEELSAIKGAELEKPEEGEEPEQEMDIDELIKAEFR  
ENEPGPIVGEELLISQKVFANFNMDDDELPELPEONLRMEEIIPVORRIEER  
QKELEKIYALPRRNCAKQISFNGSERHRSRRYSQSDSISERKRPKRGRPTI  
PRENLIGFSDAELRRFKSYKFKGGLERDALRAELVDKSETDLRLRGELVHNGC  
IKALNDSFVQERAAGRLGKVGPTFRISGVQVNAKLVISHEELAPLHKSIPSDPEE  
RRRYVPCHTKAHFDIDCKEDDSNLLIGIYGYGSMEMIKMDPLSLYQKILPDD  
PDKK"

BASE COUNT 947 a 464 c 613 g 730 t  
ORIGIN

Query Match 1.2%; Score 81; DB 5; Length 2754;

Best Local Similarity 100.0%; Pred. No. 3.6e-29; Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1695 GCTGATTCATGTCGAAGAATAGTTGATTCCTGCAGATGAATGGCTGGTAAA 1754  
|||||  
DB 343 GCTGATTCATGTCGAAGAATAGTTGATTCCTGCAGATGAATGGCTGGTAAA 402  
|||||

QY 1755 ACATTAACAACATTTCTTT 1775  
|||||

DB 403 ACATTAACAACATTTCTTT 423  
|||||

Mon Aug 5 11:51:35 2002

us-08-973-363-10.oli.rge

Page 15

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 01:50:11 ; Search time 949.74 Seconds  
(without alignments)  
11945.765 Million cell updates/sec

Title: US-08-973-363-10

Perfect score: 6608

Sequence: 1 CGGGCTGCGGACGAGAGGC.....AACTAAACCAAAAAAAC 6608

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size: 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N\_Geneseq\_032802:\*

1: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/hold-geneeq/geneeqn-emb1/NA2002.DAT.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6608	100.0	6608	18	AAT42751
2	103	1.6	153	18	AAT42757
3	59	0.9	1316	18	AAT42754
4	49	0.7	4222	23	AA585967
5	47	0.6	153	18	AAT42758
6	38	0.5	153	18	AAT42759
7	30	0.5	212	21	AA063280
8	30	0.5	6237	24	AA063280
9	29	0.4	6191	24	ABU33570

C	10	27	0.4	5474	24	ABL33795	Human immune system
C	11	26	0.4	300	20	AA213346	Human gene expression
C	12	26	0.4	2785	22	AAK75594	Human immune/hematopoietic
C	13	26	0.4	5919	22	AA564349	Tumour suppressor
C	14	26	0.4	5919	24	ABL32732	Human immune system
C	15	26	0.4	5919	24	AA561168	Human gene regulator
C	16	26	0.4	40862	24	ABL34073	Human immune system
C	17	25	0.4	140	21	AA013041	Human secreted protein
C	18	25	0.4	140	21	AA027949	Human secreted protein
C	19	25	0.4	153	18	AAT42756	Mouse CHD-1 gene (chromatin remodeling factor)
C	20	25	0.4	307	22	AAK58905	Human immune/hematopoietic
C	21	25	0.4	329	21	AA442060	Human secreted protein
C	22	25	0.4	331	20	AAV88585	EST clone FM375
C	23	25	0.4	332	22	ABAI1459	Human nervous system
C	24	25	0.4	412	22	AA192280	Human polynucleotide
C	25	0.4	414	22	AA534304	Human cDNA encoding	
C	26	0.4	423	21	AA030813	Human secreted protein	
C	27	0.4	423	21	AA030813	Human secreted protein	
C	28	25	0.4	550	21	AA238845	Human pax-6 conserved
C	29	25	0.4	596	22	AA123125	Human breast cancer
C	30	25	0.4	693	22	AA124263	Human breast cancer
C	31	25	0.4	700	22	AAH92867	Human inflammatory
C	32	25	0.4	728	22	AAK82985	Human immune/hematopoietic
C	33	25	0.4	770	21	AA044102	Human Adipodopsin
C	34	25	0.4	846	22	AAAI5364	Human nervous system
C	35	25	0.4	972	21	AA074431	Human secreted protein
C	36	25	0.4	1157	22	AA063808	Human secreted protein
C	37	25	0.4	1177	24	AA062678	Human secreted protein
C	38	25	0.4	1300	21	AA081109	Human secreted protein
C	39	25	0.4	1311	19	AAV95280	Altered telomere repeat
C	40	25	0.4	1352	21	AA079786	Human secreted protein
C	41	25	0.4	2099	20	AAV80615	Kidney injury associated
C	42	25	0.4	2228	22	AAK4901	Human full-length
C	43	25	0.4	2407	22	AA01890	Human secreted protein
C	44	25	0.4	2935	22	AA19006	Human secretory protein
C	45	25	0.4	2935	22	AA163356	Human kidney related
C	46	25	0.4	3379	22	AAK94754	Human full-length

#### ALIGNMENTS

RESULT 1		
AA042751	standard; cDNA; 6608 BP.	
XX	AA042751:	
AC	12-MAR-1997 (first entry)	
XX		
DT	12-MAR-1997 (first entry)	
XX		
DE	Chicken CHD-1A gene.	
XX		
KW	Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;	
KW	CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.	
XX		
OS	Gallus sp.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	228..5390
XX		/*tag- a
XX	W069639505-A1.	
PN		
PD	12-DEC-1996.	
XX		
PF	05-JUN-1996;	96WC-GB01341.
XX		
PR	06-JUN-1995;	95GB-0011439.
XX		
PA	(ISIS-) ISIS INNOVATION LTD.	
XX		
FI	Griffiths R, Tiwart B;	
XX	WPI; 1997-043127/04.	
DR		





```
Db 1801 aacgtatagcccttctctcgcgctgccacttccacttgcacatcttgcgaagag 1860
Oy 1861 AGATTCAAACTGGGCTCCTCAGATGATGCTGTGATTACTTAGAGATATTAACTA 1920
Db 1861 agattcaaaacttggctcctcagaatgactgtgttactttagagataataactgta 1920
Oy 1921 GAATATGATTAAGACATCATGAATGATGATCCAGACACTTAACATTMAAGTTTACA 1980
Db 1921 gaaataagataagactcatgaatgactgcatccacagataaagattaa 1980
Oy 1981 TACTTCGACGATATGATGAATTTTACTGAGAGATTAATCTTCCTGGTGCCTCAAT 2040
Db 1981 tacttcgacgacataatgaaatttacttgaggaatgaagtcattccttggctgctca 2040
Oy 2041 GGGCATTCATAGAGATGTTGATGAAGCTCATGTTTAAAAATGATGACTCTCTGTG 2100
Db 2041 gggcatcatagaggttgaagactcatcgtttaaaaaatgattgactctctctgta 2100
Oy 2101 GGACTTTAATAGCTTTAATAGTCACATCGACTTCTGATTCATGACGACCCACATG 2160
Db 2101 ggaacttaataagactttaaagctccacacatgactctgtattacttgcgaaccac 2160
Oy 2161 ATTCCCTCAAGAGCTGCTGCTTGTGATTTGATTCATATGCCGAAAAATTTGCTCT 2220
Db 2161 attccctcaaaagctgctgcttctgtgattcaatcaatgcgaataaattctctcc 2220
Oy 2221 GCGAAGATTTTGAAGAGCATGCGCAAGAGAGAGATATGTTATGCACTTTTACA 2280
Db 2221 gggaaagatttgaagagcatgcgcaagagagatattgatttgcactttcaca 2280
Oy 2281 AAGACTTGAAACCTTTTACTAAGAAAGCTTAAAAAGATGTAGAAAAGCTTTTACCT 2340
Db 2281 aagacttgaaaccttttactaagaaagcttlaaaaaagattgtagaagcttcta 2340
Oy 2341 CTAAAGTTGAACAATTTCTGAGATGAGAAAGATGATTCAGACAGCAATATTACA 2400
Db 2341 ctaaagttgacaatcttctgagatgagaaatgagatcctgacaaagcaattaca 2400
Oy 2401 GGATTTTAAGACGAATTTATAAGCCTCTCAGTAAAGCTTCAAAAGGATACCTGAG 2460
Db 2401 ggaattttaagacaatttataagcctctcagtaaaggttcaaaaggttacctag 2460
Oy 2461 TTCTGACATTTATGATGATTAAGAGGTTGTAACCATGCTGCTCATTAAGCCAC 2520
Db 2461 ttctgacatttatgataagagctcctcagtaaaggttcaaaaggttacctag 2520
Oy 2521 CAGATGATTAATGAATCTATATATAACAGAGGCTTACAGCATTTGATAGCTAGC 2580
Db 2521 cagatgatatgaattctataataacagaggttctacagcatttgatagctagcag 2580
Oy 2581 GGAACCTAATCTTCTTGAACAAGCTACTGATTCGTCTGCGAAGCGTGGCACAAG 2640
Db 2581 ggaacctaatcttcttctgaagaagctactgactcgtcgagaaagctggcaag 2640
Oy 2641 TGAATTTTCTCAGATGATGATGATCTGACATCTCTGCAAGAAATCTGAAGTTC 2700
Db 2641 tgaaatttctcagatgagatgagatgctggaactcctcagaaatcatctgaag 2700
Oy 2701 AGTTTCCCTTCAGAGACTTGATGATCAATTAAGGGAATTTGAGGAAGCAAGCAT 2760
Db 2701 agtttcccttcagagacttgatgataaagggaaattgaaagaaagcaagcag 2760
Oy 2761 ATCATTTCATGACAGAGATCAAGAGATTTCTGTTTTTACTGCTACAGAGCTGAG 2820
Db 2761 atcaattcaagagaagatcaagagattctgctgttttacttctacaagaagctgag 2820
Oy 2821 GATTAGTATTAACTTGGGATCTGCTGACACTGTGATTTTGTGATTCGATCGAATC 2880
Db 2821 gattaggtattaaacttgggactctgacactgttagtatttgcactgcagctgaa 2880
Oy 2881 CACAGAAATGATTCGACAGCAGGAGAGCTCATGAAATTTGACAGAAACAGGTTA 2940
Db 2881 cacaagaatgatctgcagagcagagcagagctcatagaatttgcacagaagaacsg 2940
Oy 2941 ATATTATTCGGCTATGTCACAAAAAGGATCAGTAGAAGACATATTTCTTGAAG 3000
Db 2941 atattatcggctatgtacaaaaagatcagatagaagaatattcttgaagaagcca 3000
Oy 3001 AGAAGATGCTGTAGACCATTTTATGTAATTCAGAGATGAGACAGACGAAAACT 3060
Db 3001 agaagatgtgtctagacacatttagtaattcagagaaatgcacagacagaanaa 3060
Oy 3061 TGCATCAGGCTTCAACCTCATCAACCTCTACACTTTTATTAAGAGATTTATCAG 3120
Db 3061 tgcatacaggttcaactccatcaagctctacaccccttlaataaagaagttatca 3120
Oy 3121 TTTTGAAGTTTGTGCTGAGAGACTCTTTAAAGACCTGAAAGGAGAAAGAGGCC 3180
Db 3121 ttttgaagtttgtgtctgagagactctttaagaaccttgaaggagaagacagag 3180
Oy 3181 AGGAAATGATATGATGAAATCTTGAAGAGAGCTGAAACCTCGGAAAAAGAGCG 3240
Db 3181 aggaatatgatatagatgaataccttgaaagagcttgaabctcgggaaaaatgag 3240
Oy 3241 CATGACTGTAGGGAATGAGACTCTTTCACAGTCAAGGTGCGAATTTTCCATATG 3300
Db 3241 catgactgttagggatgagactctttcacagtccaagctcgaagctcttccaaat 3300
Oy 3301 ATGAGATGATATTTGACTTGGAACCAAGAAATTTCAAGAAATTTGGAGAAATC 3360
Db 3301 atgagatgatatttgacttggaaaccaaaagaaattcaagaaatttggagaaatc 3360
Oy 3361 CAGATCCCAAGAGAGATTAAGAGAGAGAAAGACAAAGAAATTTGAAGAAATAT 3420
Db 3361 cagaatcccaagagagatagaagagagagagagacaaaagaacttgaagaatat 3420
Oy 3421 ACATGCTCCGAGGATGAGAAACTGTGCAAAACAGATCAGCTTTAATGGAGTGA 3480
Db 3421 acatgctccgagagatgagaaactgtgcaaaaacagatcagctttaatggagtg 3480
Oy 3481 GAGCGATGAGAGCAGAAATATTCTGATCTGATAGTACTCCATCAGAAAGAAAC 3540
Db 3481 gagcgatgagagcagaaatatttctgactgatatgactccatcagaaagaaac 3540
Oy 3541 GGCCAAAAAAGCGTGAAGACCTCGAACATTTCTCGAAGAAATTTGAAGATTAG 3600
Db 3541 ggcctaaaaagcgtggaagacctcgaaacctctcgaagaataattaaagattag 3600
Oy 3601 ATCAGAGATCAGCGCTTTATCAGAGATTTACAAGAAATTTGTGGGCTCTGAA 3660
Db 3601 atcagagatcagcgctttatcagaagttacaagaatttgttggccctcgaagag 3660
Oy 3661 TAGATCTGTAGCTAGAGATGCTGAACTGTTGATTAATCTGAGACAGACTTGA 3720
Db 3661 tagatctgtagctagaagatgctgaaactggttgaataaactcgagacgaactga 3720
Oy 3721 TGGGTAACTTGTATATATGATGATTAAGGCTTTAAAGGACAAATTCATCTG 3780
Db 3721 tgggtaaacttgtatataatgatgatataagctttaaaggaacaattcatctgag 3780
Oy 3781 AAAGACAGAGAGTGAAGCTTGGGAAAGTTAAAGCCCAAGCTTTCAGATTCAG 3840
Db 3781 aaagacagagagtagaacttgggaaagtlaaagcccaagcttctgaatctcga 3840
Oy 3841 AGGTGATGCAAAACTAGTATCTCTCAGAAAGAGAGCTGGCACCACTGCACAAT 3900
Db 3841 aggtgaatgcaaaactagatctctcagaaagaagctggcaacactgcacaaatca 3900
Oy 3901 TTCTTCAGATTCAGAAAGAAAGAAAGATATGTCATCCATGCCACACAGCTGCT 3960
Db 3901 ttcttcagattcagaaagaaagaaatattgtcatcccaagcacaagaagctgct 3960
Oy 3961 ACTTCGATATGATTTGGGTAAGAAAGATGATTCGAATCTGTTAGTGGCATATGA 4020
Db 3961 acttcgatatagatitgggtlaaagaagatgattccaactctgttagtgatctatga 4020
```

QY 4021 ATGCGTATGCGACGTGGGAAATGATTTAAATGATTCAGATCTCAGCTTAAACAGAGA 4080  
|||||  
Db 4021 atgctatgagcgttggaaatgataaaatgataccagatctcagcttaacacagaga 4080  
QY 4081 TTTTACCTGATGATCCAGACAGAAACCCAGGCAAAAGCAGCTACAGACCCGTGCAGACT 4140  
|||||  
Db 4081 tttaactgtagatccagacagaacccagcaagcagctacagcccgctgagact 4140  
QY 4141 ACCCATTTAAATTTCTGATTAAGACCTTGCAGAAAGAGCGCAAAAGCTTCTGTGTG 4200  
|||||  
Db 4141 acccatlaaatctaacgaaataaaccttcgaagaagagcacaagagcttgcgtg 4200  
QY 4201 CAGGCAATTTCCAGAGAGAGAGAGCAAGAAATAGAAAGATAGATGAAGCTTCAAAAA 4260  
|||||  
Db 4201 caggcaatctccaagagagaagagcaagaataaagaataaagatgaaagcttcaaaaa 4260  
QY 4261 TAAAGAAGAAATTAAGAGTATTTCTTCCACCAACCTTCAGAAAATCTGATGAAGATG 4320  
|||||  
Db 4261 taaagaagaataaagatgattcttccacacacccctcagaanaatctgtagaagt 4320  
QY 4321 ATGAGGAGGAGATACAGGTAAGTAATGAATGAATCTGAATAAGAAATCTTAA 4380  
|||||  
Db 4321 atgagagagagagatacaagaatgaaatgaaatctgaaatacaagaanaatctlaaa 4380  
QY 4381 AAATTCATTTGCTGGATACTCCAGTTCATATTACTGCAACCCAGTGAACCAAGTTCATVCT 4440  
|||||  
Db 4381 aaatccatctgctgtagatctcagctcagttcatctactgcaacagtagaacagcttctatct 4440  
QY 4441 CAGAGAATCTGGAAGAACTCCATCAGAAAGACATTAGTGTGTGCAAGAAAGATGAGGC 4500  
|||||  
Db 4441 cagaagaatctggaagaatcccatcagaagaatctgtagtctgcaagaagaatgagc 4500  
QY 4501 CTGCAAGAGAGAGCTGCTGAACAGCTGATAGACAGAGAAAGGCTTTCTGAAGGGAGC 4560  
|||||  
Db 4501 ctgtcaaaagcagcagctgaacagctgagtagaacagaagaagagccttctgaaagagc 4560  
QY 4561 AGCTGGAACATCTAGGCACTGTCTTAATCAAAATTTGGGATCAGATTACAGAAATGCCGTA 4620  
|||||  
Db 4561 agctggaacatactagcagagcttaatacaaaatctggagatcacatcagaatctgcga 4620  
QY 4621 AGGAGTACACAAATCCGAGCAATTAATAACAGTGGAGGAAAAATTTGTGGATTTTGTGT 4680  
|||||  
Db 4621 aggagtagacaaaatcccgagaaataaaacagtaggagaaaatctgtggttctgtgt 4680  
QY 4681 CCAAGTTTACAGAAATTTGATGCCAGAAAGCTGCACAAACTCTACAAACATGCAATCAAAA 4740  
|||||  
Db 4681 ccaagtttaagaatttgatgagcagaagaagctgcacaaactctacaacatgcaatcaaaa 4740  
QY 4741 AGGCGCAAGAGCTTCAGCAACACAAATGACCAAAACATTTAGCAGCAATGTGAATACATG 4800  
|||||  
Db 4741 agcgccaagagcttcagcaacacaaatgaccaaactatgacgaatctgtaatacacatg 4800  
QY 4801 TAAACAGAAATCCAGATGTGGAAAGACTGGAAGAGACTACCAAAACATGATATGTA 4860  
|||||  
Db 4801 taaacaaaataccagatgtagaagaagactggaagactacaacacatgatagtatgtagca 4860  
QY 4861 GGGACAGTATTTCTTCTGATAGACATTTATCACAATACCATGATCACAAGACAGC 4920  
|||||  
Db 4861 gggacagtattctctctgtagatgacattatcaacatacatgatacacaagaagcggc 4920  
QY 4921 ATTCAGGAGATGCTTACAGAAAAAGTACTCAGGAAAAGGCCATATTACGCTTCACTA 4980  
|||||  
Db 4921 atcagggagatgcttacaagaagaagtactcagaaagccatatctcagccttcagta 4980  
QY 4981 ATGGAAGAAAGATCAGAGAGCTGGGATCATCTTCAACAGAGAGAGAGATACTACAGTATA 5040  
|||||  
Db 4981 atggaaaaagatcacagagagacgggagtagcaactacaacagagacagatactaccagtagata 5040  
QY 5041 GTAACAATAGAAAATTAGATGACACAGAGAGAGAGACAGAGTCAAACTGGAAGGAA 5100  
|||||  
Db 5041 gtaaacatagaaaattagatgagacacagagcgagagacacagagctcaaaccttgaaagaa 5100

QY 5101 ACTTAAAGACAGCGGGGATTCAGATCACCCGCTCCATTTAGACACAGATACACT 5160  
|||||  
Db 5101 acttaaaagacagcggggtatctcagataccacgcgtcccatctcagaccacagataact 5160  
QY 5161 CAGATCACCGTTCCACTTCAGAAATACAGCCATCATTAATCTTGCAGAGATTTATGATACC 5220  
|||||  
Db 5161 cagatcaccgcttccactctcagatacagccatcaataatctctcgagatlatagatacc 5220  
QY 5221 ACTCAGACTGGCAANTGGACACAGAGCTTCTGTAGTGGCGCGAGGTCACCATACATC 5280  
|||||  
Db 5221 actcagactggcaaatggacacagagcttctgtagtggcgcgaggtcacacataagtc 5280  
QY 5281 AGAGGTCTCCTTAATGTTCAAGATCTCCCTAGACACAGATCTCATTTGAAACATCAT 5340  
|||||  
Db 5281 agaggtctccttattggttcaagatctccctcaggacaagcttccatcttgaacatccat 5340  
QY 5341 CAGATCACAAAGTACACTTACATCATATGAGATAGCCGGAACACATTAACAAAGACTGA 5400  
|||||  
Db 5341 cagatcacaaaagtagacacttgaacataatgagtagccggaaagacatacaaaagactga 5400  
QY 5401 CATTTCTGACCTTTCTTTTAAAGCCATATACAGTAAACTAACAGTAAATGGCTTACAT 5460  
|||||  
Db 5401 catlttctggaactcttctttagccataacagtagaactaacacagtaattgacctacat 5460  
QY 5461 GACTTGAAGATATGACTGATATTTCTATCATAGCAGATATGTTCTACTTCTTCCAGA 5520  
|||||  
Db 5461 gacttgaagaatagagactgagatattctatcagtagagagtagtcttactcttccagga 5520  
QY 5521 TGCAAGGTCTTATTTCCACACAGAAAGAAATATTTTGTATTTAAAGTTTATCTCTCAC 5580  
|||||  
Db 5521 tgcgaaggtcttattttccacacagaaagaaatatttttgtattttaaagttatgctgcac 5580  
QY 5581 TGTGCTGCAAAATGTTGTGACCTTTTAAAGAAATGGAAGATGTTTACTTTTACAGG 5640  
|||||  
Db 5581 tgtgctgcaaaatgttgtgacactttttaaagaaatggaagatggttacttcttaacgg 5640  
QY 5641 GACCTCAACACTGCCCCCTTTCAGACTGAGATCTTAAACTCTTCATGTCAAGTGG 5700  
|||||  
Db 5641 gacctcaaacactgcccccttccagactgagatcttctataaacccttcatctgcaaaagt 5700  
QY 5701 TTCTAGGCTGACACAGATTAATTAATGTTGTAATGAACACTTAACACTGACCTGTG 5760  
|||||  
Db 5701 ttctagagctgaaacagatlaaatlaatgcttgaataaagaaactlaaacctgcgcgtg 5760  
QY 5761 CTTATGTTTTCAGAAAGATGGGGATTTATTTGTTTATTTCTTGTGTAGAGAACTCTC 5820  
|||||  
Db 5761 cttaatgcttcaagaaagaaatgagggatttatcttcttacttctctgttagagaaacttc 5820  
QY 5821 AAGGACTTTGTTCACCTTCCAAAGCTACTTGTATGATGTGACCTGCGACACCTTGCC 5880  
|||||  
Db 5821 aaggacttcttcaacttccaaagctacttgttatacttgaacttgcagaccacttgc 5880  
QY 5881 GCTTTTATCCAGAGTGTGAATATTTAAATTTCTTACTCTCAGTTGTAATATCCAGGA 5940  
|||||  
Db 5881 gcttttataccaagacttgaatatttaaatctgtacttaaaagttgtaaaatgacccagga 5940  
QY 5941 TTTCTCCTGTTTGTGATCAGTTATTAATGCCTTTATGAAACAAACAAACAAAAAA 6000  
|||||  
Db 5941 ttctcctgtttagatcagtagtataatgaccttctttagaaacaaaacaaacaaaataaa 6000  
QY 6001 CAATTTAAAAAAAACACAAACAAACCAACAAATGCGTGTAAATTTTGTAAATTAAT 6060  
|||||  
Db 6001 caatttaaaaaaaacacaaacaaacaaacaaatgctgttaaatltgttaaatlaatt 6060  
QY 6061 AAATGAGCTTTTTCGTCAGGCTTTTGGCGTGTCTTCCCAACAACTCAGGCT 6120  
|||||  
Db 6061 aaatgagcttcttccgtcagagcttcttctgtgcgtctcctccccaacactcagagcct 6120  
QY 6121 TCTTTTCAACAAGTCAATATCTTACATGTTTAAATATATTCGATGGAATCACAAT 6180  
|||||  
Db 6121 tcttttcaacaagctcagtagataacttaacatgltttaaataaatctcgaatcagaaat 6180  
QY 6181 GTAAAAATGGGGAAGGAATATTTTATTCATTTAGTGTCTCTTTTATTTGATACTTT 6240



Query Match 0.9%; Score 59; DB 18; Length 1316;  
Best Local Similarity 100.0%; Pred. No. 6.3e-16;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2343 ATTATGCGGTACACAAAAGATCAGTAGAAGATATCTTGAAGGCCAAGAA 3001  
|||||  
Db 1 attatcgctagtcacaaagatcagtagaagaagatattcttgaagaagcaagaa 59

## RESULT 4

AAS85967/C  
ID AAS85967 standard; cDNA; 4222 BP.

AC AAS85967;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #21771.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Dmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG21780.

PT New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity -

PS Claim 1; SEQ ID No 21771; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at [ftp.wipo.int/pub/publ/published\\_pct\\_sequences](http://wipo.int/pub/publ/published_pct_sequences).

SO Sequence 4222 BP; 1040 A; 1041 C; 883 G; 1258 T; 0 other;

Query Match 0.7%; Score 49; DB 23; Length 4222;  
Best Local Similarity 100.0%; Pred. No. 1.7e-11;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3174 GAGCCCCAGAGAAATGATATGATGAATCTTGAAGAGAGCTGAACCTC 3222  
|||||  
Db 1765 GAGCCCCAGAGAAATGATATGATGAATCTTGAAGAGAGCTGAACCTC 1717

## RESULT 5

AAT42758  
ID AAT42758 standard; DNA; 153 BP.

AC AAT42758;

DT 12-MAR-1997 (first entry)

DE Chick CHD-W gene fragment.

KM Bird; sex determination; chromodomain-helicase-DNA binding 1;

KW CHD-1A; CHD-W; W chromosome; ss.

OS Gallus sp.

PN Key Location/Qualifiers

FT misc\_difference 52..81

FT /\*tag- a

FT "bases 52-81 are a repeat of bases 22-51  
and are ignored in the translated amino  
acid sequence given in Fig 3"

PN WO9639505-A1.

PD 12-DEC-1996.

PF 05-JUN-1996; 96WO-GB01341.

PR 06-JUN-1995; 95GB-0011439.

PA (ISIS-) ISIS INNOVATION LTD.

PI Griffiths R, Tiwari B;

DR WPI; 1997-043127/04.

DR P-PSDB; AAW08148.

PT Avian chromodomain-helicase-DNA binding genes determine sex in  
birds - used for sex determin. and to control sex of progeny  
XX  
XX  
XX Claim 8; Fig 3; 76pp; English.

XX Bases 3855-3977 (AAT42756) of the mouse CHD-1 gene show homology  
XX to portions of the chicken CHD-1A (A = Avian) gene (AAT42757)  
XX and chicken CHD-W (W refers to the W chromosome) gene (AAT42758) and  
XX and the great tit CHD-W gene (AAT42759). Translated amino acid  
XX sequences of this region are provided in AAW08146-49. The CHD-1A  
XX (see also AAT42751) and CHD-W (see also AAT42754-55) genes determine  
XX sex in birds and can be used to identify the sex of an embryo,  
XX foetus etc. and to manipulate the sex of progeny.

SO Sequence 153 BP; 56 A; 36 C; 31 G; 30 T; 0 other;

Query Match 0.7%; Score 47; DB 18; Length 153;  
Best Local Similarity 100.0%; Pred. No. 1.5e-10;  
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4140 TACCTCATTAATTAATGATAAAGACCTTGCAAGAAAGACGACA 4186  
|||||  
Db 91 taactcataattactcgaataaagaccttgcaagaagaagacaca 137

## RESULT 6



XX 03-JAN-2002.  
PD  
XX  
XX 02-JUL-2001; 2001WO-EP07537.  
PF  
XX 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPIC-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
DR  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX  
PS Claim 1; SEQ ID NO 331; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 6237 BP; 1614 A; 77 C; 1263 G; 3283 T; 0 other;

Query Match 0.5%; Score 30; DB 24; Length 6237;  
Best Local Similarity 100.0%; Pred. No. 0.0051;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5979 AACCAACCAACCAACCAACCAATTAA 6008  
DB 2795 AACCAACCAACCAACCAACCAATTAA 2766  
|||||

RESULT 9  
ABL3370/C  
ID ABL3370 standard; DNA; 6191 BP.  
XX

AC ABL33370;

XX 26-MAR-2002 (first entry)  
DT  
XX

DE Human immune system associated gene SEQ ID NO: 1343.  
XX

KW Human: immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosinatic; noctropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antineumatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX

OS Homo sapiens.  
XX

PN WO200200928-A2.  
XX

PD 03-JAN-2002.  
XX

PF 02-JUL-2001; 2001WO-EP07537.  
XX

PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX

PA (EPIC-) EPIGENOMICS AG.  
XX

PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
DR  
XX  
XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX  
PS Claim 1; SEQ ID NO 1343; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 6191 BP; 1678 A; 92 C; 1597 G; 2824 T; 0 other;

Query Match 0.4%; Score 29; DB 24; Length 6191;  
Best Local Similarity 100.0%; Pred. No. 0.014;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5979 AACCAACCAACCAACCAACCAATTAA 6007  
DB 328 AACCAACCAACCAACCAACCAATTAA 300  
|||||

RESULT 10  
ABL33795/C  
ID ABL33795 standard; DNA; 5474 BP.  
XX

AC ABL33795;

XX 26-MAR-2002 (first entry)  
DT  
XX

DE Human immune system associated gene SEQ ID NO: 1768.  
XX

KW Human: immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosinatic; noctropic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antineumatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX

OS Homo sapiens.  
XX

PN WO200200928-A2.  
XX

PD 03-JAN-2002.  
XX

PF 02-JUL-2001; 2001WO-EP07537.  
XX

PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX

PA (EPIC-) EPIGENOMICS AG.  
XX

PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2002-130909/17.  
DR  
XX

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -  
XX  
XX  
PS Claim 1; SEQ ID NO 1768; 32pp + Sequence Listing; German.  
XX



PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 01-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249287.  
PR 17-NOV-2000; 2000US-0249289.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

## (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -

PS Disclosure; SEQ ID NO 30406; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX  
SQ Sequence 2785 BP; 840 A; 572 C; 594 G; 779 T; 0 other;

Query Match 0.48; Score 26; DB 22; Length 2785;  
Best Local Similarity 100.0%; Pred. No. 0.32;





RESULT 15  
AAS61168/c  
ID AAS61168 standard; DNA; 5919 BP.  
XX  
AC AAS61168;  
XX  
DT 29-JAN-2002 (first entry)  
XX

Human gene regulation-associated gene oligonucleotide #123.

XX  
DE Human: Gene regulation-associated gene; severe combined immunodeficiency;  
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;  
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;  
KW renal disease; Preeclampsia; cardiac allograft vascular disease;  
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;  
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;  
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.  
XX

OS Homo sapiens.

PN MO200177375-A2.

XX 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP03968.

XX 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Plegenbrock C, Berlin K;

PI WPI; 2002-017470/02.

XX New nucleic acid sequences from chemically modified genes associated  
PT with gene regulation, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency  
PT disease

PS Claim 1; SEQ ID No 127; 26pp; English.

XX  
CC The invention relates to 224 nucleic acid sequences comprising at least  
CC 18 bases of a chemically pretreated gene associated with gene regulation  
CC selected from 43 known genes (or complementary sequences). The  
CC chemical pretreatment converts cytosine bases unmethylated at the  
CC 5-position to uracil or another base with hybridisation behaviour  
CC dissimilar to cytosine, to enable analysis of cytosine methylations.  
CC The DNA sequences, oligomers (or sets/arrays) and method are  
CC useful in the diagnosis of diseases (or predisposition to diseases)  
CC associated with gene regulation and in therapy of such diseases, by  
CC enabling analysis of the cytosine methylation patterns of such genes,  
CC kits are provided. They are especially useful in diagnosis  
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac  
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,  
CC preeclampsia, graft versus-host disease. The present sequence is a  
CC sequence included in the sequence data for this specification and is  
CC associated with the human gene regulation-associated genes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5919 BP; 1421 A; 69 C; 1641 G; 2788 T; 0 other;

OY 5979 AACCAACAAACAAACAAACAAAT 6004  
|||  
Db 3118 AACCAACAAACAAACAAACAAAT 3093


Search completed: August 3, 2002, 01:52:38  
Job time: 32343 sec

Query Match 0.4%; Score 26; DB 24; Length 5919;  
Best Local Similarity 100.0%; Pred. No. 0.31;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Mon Aug 5 11:51:35 2002

us-08-973-363-10.oli.rng

Page 13



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:40 ; Search time 7016.61 Seconds

(without alignments)  
12710.964 Million cell updates/sec

Title: US-08-973-363-10

Perfect score: 6608

Sequence: 1 CGGCTCGCGACGAGAGCGC.....AACTAAACCAAAAAAAC 6608

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 674847542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : EST.\*

1: em\_estha:\*  
2: em\_esthm:\*  
3: em\_estln:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	443	6.7	597	10	BM486590
2	394	6.0	648	10	BM491730
3	87	1.3	476	10	BE479788
4	82	1.2	234	9	AM075773
5	82	1.2	252	9	AA824571
6	82	1.2	296	9	AA480460
7	82	1.2	352	10	H80753
8	82	1.2	354	9	AA483709
9	82	1.2	399	10	H99726
10	82	1.2	400	10	H99726
11	82	1.2	408	9	AM591975
12	82	1.2	419	9	AA442118
13	82	1.2	434	9	AA609977
14	82	1.2	436	10	N20387
15	82	1.2	466	9	AT088996
16	82	1.2	466	9	BE048526
17	82	1.2	490	9	AA599181

18	82	1.2	505	9	AI298448
19	82	1.2	642	10	BE973688
20	82	1.2	789	10	BE613057
21	82	1.2	802	10	BC391166
22	82	1.2	1280	10	BM015816
23	79	1.2	945	10	BE535223
24	74	1.1	422	10	RA4604
25	67	1.0	545	10	BF559370
26	67	1.0	543	9	AI233881
27	67	1.0	551	9	AI409754
28	67	1.0	566	9	AI176714
29	67	1.0	597	9	AI406778
30	67	1.0	613	9	AA943664
31	67	1.0	637	9	BE107423
32	66	1.0	633	9	AU045087
33	66	1.0	353	9	AU046169
34	66	1.0	364	10	BC071866
35	66	1.0	442	9	AA273291
36	66	1.0	486	9	AA116690
37	66	1.0	495	10	BE447654
38	66	1.0	531	10	BI794545
39	66	1.0	1346	11	AK018451
40	63	1.0	238	10	BF704506
41	63	1.0	343	9	AA885644
42	63	1.0	401	10	BE550103
43	63	1.0	509	10	BE813029
44	63	1.0	539	10	BE812861
45	63	1.0	549	10	BE813014

## ALIGNMENTS

RESULT 1  
LOCUS BM486590 597 bp mRNA linear EST 07-FEB-2002  
DEFINITION pgm2n.pk002.b16 Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus cDNA clone pgm2n.pk002.b16 5' similar to gblAAC60282.1 (AF004337) chromo-helicase-DNA-binding on the z chromosome protein [Gallus gallus], mRNA sequence.  
ACCESSION BM486590  
VERSION BM486590.1 GI:18607520  
KEYWORDS EST.  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 597)  
Cognburn, L.A. and Monsonego-Ornan, E.  
ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library, USDA/IFARS Animal Genome Project

JOURNAL Unpublished (2002)  
COMMENT Contact: Larry A. Cognburn  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1335  
Fax: 302-831-2822  
Email: cognburn@udel.edu, www.chickest.udel.edu.

FEATURES  
source location/Qualifiers  
1..597  
/organism="Gallus gallus"  
/strain="Commercial broiler and Ottawa Res. Centre  
Strains 90 & 21"  
/DB\_xref="taxon:9031"  
/clone="pgm2n.pk002.b16"  
/clone\_lib="Normalized Chicken Breast Muscle, Leg Muscle,  
and Epiphyseal Growth Plate cDNA library (pgm2n)"  
/sex="Male and Female"  
/tissue-type="Breast muscle, Leg muscle and epiphyseal  
growth plate"

```

/del stage="Breast_leg:Embrryo(d19);post-hatch(1d,1,3,5,7,9
,11 weeks):growth_plate(1d,7d,14d post-hatch)"
/lab host="E. coli EMDH10B"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue (embryonic
muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth
plate 33.3% of the final RNA pool). Single pass sequencing
from 5'-end"
BASE COUNT      186 a      126 c      106 g      179 t
ORIGIN

```

Query Match	6.7%	Score 443	DB 10	Length 597
Best Local Similarity	99.8%	Pred. No. 1.le-173		
Matches 563	Conservative	0	Mismatches 0	Indels 1
				Gaps 1
QY 5172	TCACCTTCAGAAATACAGGCATCATATAAATCTTCGAGAGATTTAGATATACCACCTACAGACTGG	5231		
Db 1	TCCACTTCAGAAATACAGGCATCATATAAATCTTCGAGAGATTTAGATATACCACCTACAGACTGG	60		
QY 5232	CAAAATGCACACAGAGCTTCTGTAGTGGCCCGAGGTCACCACTGATCAGAGCTCTCT	5291		
Db 61	CAAAATGCACACAGAGCTTCTGTAGTGGCCCGAGGTCACCACTGATCAGAGCTCTCT	120		
QY 5292	TATGGTTCAAGATCTCCCTTAGAGACAGATCTCCATTTGAAACACTCATCAGATACAAA	5351		
Db 121	TATGGTTCAAGATCTCCCTTAGAGACAGATCTCCATTTGAAACACTCATCAGATACAAA	180		
QY 5352	AGTACACCTGAAACATACATGAGTGGCCGGAAGACATTAACAAAGACATGATCTCTGGA	5411		
Db 181	AGTACACCTGAAACATACATGAGTGGCCGGAAGACATTAACAAAGACATGATCTCTGGA	240		
QY 5412	CCCTTCCTTTTACCATATACAGTAAACATACACAGATTAATGGCTTACATGATGAGTAAAG	5471		
Db 241	CCCTTCCTTTTACCATATACAGTAAACATACACAGATTAATGGCTTACATGATGAGTAAAG	300		
QY 5472	TATGACACGTGATATCTCATCACTAGCAGTATTTGTTACTTCTTCCAGATGCAAGGTCTA	5531		
Db 301	TATGACACGTGATATCTCATCACTAGCAGTATTTGTTACTTCTTCCAGATGCAAGGTCTA	360		
QY 5532	TATATCCCAACAGAGAAAAATATTTTGTATTTAAAGTTTATGCTGCACTGGCTGCAAA	5591		
Db 361	TATATCCCAACAGAGAAAAATATTTTGTATTTAAAGTTTATGCTGCACTGGCTGCAAA	420		
QY 5592	TGTTGTGGCAGCTTTTTTTTTTAAGAAATGGAAGATGTTTACTTTTACAGGAGCCTCAACAC	5651		
Db 421	TGTTGTGGCAGC-TTTTTTTTTAAGAAATGGAAGATGTTTACTTTTACAGGAGCCTCAACAC	479		
QY 5652	TGCCCTTTTCAGACCTGAGATCTTACTATTAACCTCTTCAATGTCAAAGGTGTTAGGCTGA	5711		
Db 480	TGCCCTTTTCAGACCTGAGATCTTACTATTAACCTCTTCAATGTCAAAGGTGTTAGGCTGA	539		
QY 5712	ACACAGATTAAATATTTGTTGTAA 5735			
Db 540	ACACAGATTAAATATTTGTTGTAA 563			
RESULT 2				
LOCUS BM491730				
DEFINITION	BM491730 648 bp mRNA linear EST 07-FEB-2002			
ACCESSION	pp2n.pk007.e18 Normalized Chicken Pituitary/Hypothalamus/Pineal			
VERSION	Library (pp2n).1 Gallus gallus cDNA clone pp2n.pk007.e18 5' similar			
KEYWORDS	to gbl AAC60282.1 (AF004397) chromo-helicase-DNA-binding on the Z			
SOURCE	chromosome protein [Gallus gallus], mRNA sequence.			
ORGANISM	BM491730.1 GI:18612661			
REFERENCE	EST.			
	Chicken.			
	Gallus gallus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Archaeosuria; Aves; Neognathae; Galliformes; Phasianidae;			
	Phasianinae; Gallus.			
	1 (bases 1 to 648)			

**AUTHORS** Porter, T.E. and Cogburn, L.A.  
**TITLE** ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA Library, USDA/IRAFs Animal Genome Project  
**JOURNAL** Unpublished (2002)  
**COMMENT** Contact: Larry A. Cogburn  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1335  
Fax: 302-831-2822  
Email: cogburn@udel.edu, www.chickest.udel.edu.  
**FEATURES** Location/Qualifiers

```

/organism="Gallus gallus"
/strain="Commercial Broiler chickens"
/db_xref="taxon:9031"
/clone="pqp2n.pk007.e18"
/clone_lib="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library (pqp2n)"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (dl2,dl4,dl9); post-hatch (1,3,5,7,
weeks)"
/lab_host="E. coli EMDH10B"
/notes="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"

```

[illegible]

RESULT 3  
LOCUS BE479788 476 bp mRNA linear EST 28-AUG-2000  
DEFINITION BE479788 164616 BARC 580V Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE479788  
VERSION BE479788.1 GI:9599321  
KEYWORDS EST.  
SOURCE Bos taurus  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 476)  
AUTHORS Sonstegard, T.S., Capuco, A.V., Van Tassell, C.P., Ashwell, M.S. and Wells, K.D.  
TITLE Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Sonstegard TS  
USDA, ARS, Beltsville Agricultural Research Center  
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA  
Tel: 301 504 8416  
Fax: 301 504 8416  
Email: tadseldpsi.barc.usda.gov  
Single pass sequencing. Bases called and alt-trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTTCACAGTACGACG  
Plate: 14 row: E column: 24  
Seq primer: ATTTAGTGACACTATAG.  
FEATURES  
source  
1..476  
Location/Qualifiers  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="BARC 580V"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV-SPORT6; Site.1: XbaI; Site.2: XhoI;  
Library made from pooled mRNA isolated from mammary  
tissues at eight physiological, developmental, and disease  
states."  
BASE COUNT 137 a 94 c 67 g 178 t  
ORIGIN  
Query Match 1.3%; Score 87; DB 10; Length 476;  
Best Local Similarity 100.0%; Pred. No. 3.3e-25;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5831 TTCACTTTCCAAAGCTACTGTTTACATGTGACACGACACCTTGGCGCTTTTCATC 5890  
DB 183 TTCACTTTCCAAAGCTACTGTTTACATGTGACACGACACCTTGGCGCTTTTCATC 242  
QY 5891 ACACGCTGAATATTAATTCGTGAC 5917  
DB 243 ACACGCTGAATATTAATTCGTGAC 269  
RESULT 4  
LOCUS AM075773 234 bp mRNA linear EST 13-OCT-1999  
DEFINITION AM075773 xabsc01.x1 NCI\_CGAP\_CML1 Homo sapiens cDNA clone IMAGE:2573568 3',  
mRNA sequence.  
ACCESSION AM075773  
VERSION AM075773.1 GI:6030771  
KEYWORDS EST.  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 234)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Elisabeth Palotta, Jonathan D. Licht, M.D.,  
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life  
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
cDNA distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/btrp/image/image.html](http://www.bio.lnl.gov/btrp/image/image.html)  
Seq primer: -400P from Glibco.  
FEATURES  
source  
1..234  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2573568"  
/clone\_lib="NCI\_CGAP\_CML1"  
/tissue\_type="myeloid cells, 18 pooled CML cases, BCR/ABL  
rearrangement positive, includes both chronic phase and  
myeloid blast crisis"  
/lab\_host="DH10B"  
/note="Organ: whole blood; Vector: PCMV-SPORT6; Site.1:  
SalI; Site.2: NotI; Cloned unidirectionally. Primer:  
Oligo dt. Library constructed by Life Technologies."  
BASE COUNT 88 a 28 c 46 g 72 t  
ORIGIN  
Query Match 1.2%; Score 82; DB 9; Length 234;  
Best Local Similarity 100.0%; Pred. No. 5.5e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 5836 TTTCACAAAGCTACTGTTTACATGTGACACGACACCTTGGCGCTTTTCATCAG 5895  
DB 198 TTTCACAAAGCTACTGTTTACATGTGACACGACACCTTGGCGCTTTTCATCAG 139  
QY 5896 CTGGAATTTTAATTCGTGAC 5917  
DB 138 CTGGAATTTTAATTCGTGAC 117  
RESULT 5  
LOCUS AA824571 252 bp mRNA linear EST 07-APR-1998  
DEFINITION AA824571 oc78h11.s1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:1355877,  
mRNA sequence.  
ACCESSION AA824571  
VERSION AA824571.1 GI:2896593  
KEYWORDS EST.  
SOURCE human  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 252)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.nih.gov  
Tissue Procurement: Louis M. Straudt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be





was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACATCTAGTGGAGCGCGCGCTTTTTCATCAGAG 5895  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pRT3 vector  
 (Pharmacia). Library constructed by Bento Soares and  
 M. Fatima Bonaldo. RNA from normal foreskin melanocytes  
 (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 123 a 56 c 68 g 102 t 3 others

Query Match 1.2% Score 82; DB 10; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-23;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5836 TTTCGAAGCTACTGTTTACATCTGTACAGTGGAGCACCCTTGGCTTTTCATCAGAG 5895  
 DB 139 TTTCGAAGCTACTGTTTACATCTGTACAGTGGAGCACCCTTGGCTTTTCATCAGAG 80  
 OY 5896 CTTGAATATTAAATCTGTAC 5917  
 DB 79 CTTGAATATTAAATCTGTAC 58

RESULT 8  
 AA483709 354 bp mRNA linear EST 15-AUG-1997  
 LOCUS ne75g04.s1 NCI\_CGAP\_Ew1 Homo sapiens cDNA clone IMAGE:910134, mRNA  
 DEFINITION sequence.  
 ACCESSION AA483709  
 VERSION AA483709.1 GI:2212522  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 354)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-r@mail.nih.gov  
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,  
 Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bprp/image/image.html  
 Seq primer: -41m3 fwd RT from Amersham.

FEATURES Location/Qualifiers

1..354  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:910134"  
 /clone\_lib="NCI\_CGAP\_Ew1"  
 /tissue\_type="Ewing's sarcoma"  
 /lab\_host="DH10B"  
 /note="Vector: PAMP10; mRNA made from Ewing's sarcoma,  
 cDNA made by oligo-dT priming. Non-directionally cloned.  
 Size-selected on agarose gel, average insert size 600 bp.  
 Reference: Krizman et al. (1996) Cancer Research  
 56:5380-5383."

BASE COUNT 124 a 58 c 53 g 119 t

ORIGIN

Query Match 1.2% Score 82; DB 9; Length 354;  
 Best Local Similarity 100.0%; Pred. No. 4.6e-23;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5836 TTTCGAAGCTACTGTTTACATCTGTACAGTGGAGCACCCTTGGCTTTTCATCAGAG 5895  
 DB 140 TTTCGAAGCTACTGTTTACATCTGTACAGTGGAGCACCCTTGGCTTTTCATCAGAG 199  
 OY 5896 CTTGAATATTAAATCTGTAC 5917  
 DB 200 CTTGAATATTAAATCTGTAC 221

RESULT 9

H99736/c 399 bp mRNA linear EST 15-DEC-1995  
 LOCUS yx27g03.s1 Soares melanocyte 2NBHM Homo sapiens cDNA clone  
 DEFINITION IMAGE:262996 3', mRNA sequence.

ACCESSION H99736  
 VERSION H99736.1 GI:1124404  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 399)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,  
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston,  
 R., Williamson, A., Wohlmann, P., and Wilson, R.  
 TITLE The Wash-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT Contact: Wilson R.  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu

Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 968 Std Error: 0.00  
 Seq primer: m13 -40 forward  
 High quality sequence stop: 400.

FEATURES Location/Qualifiers

1..399  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:262996"  
 /clone\_lib="Soares melanocyte 2NBHM"  
 /sex="Male"  
 /tissue\_type="melanocyte"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Vector: pRT3D (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI; Site 3: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dT) primer [5',  
 TGTTCACATCTAGTGGAGCGCGCGCTTTTTCATCAGAG 5895  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pRT3 vector  
 (Pharmacia). Library constructed by Bento Soares and  
 M. Fatima Bonaldo. RNA from normal foreskin melanocytes  
 (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 140 a 60 c 73 g 126 t

ORIGIN

Query Match 1.2% Score 82; DB 10; Length 399;  
 Best Local Similarity 100.0%; Pred. No. 4.3e-23;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5836 TTTCGAAGCTACTGTTTACATCTGTACAGTGGAGCACCCTTGGCTTTTCATCAGAG 5895  
 DB 191 TTTCGAAGCTACTGTTTACATCTGTACAGTGGAGCACCCTTGGCTTTTCATCAGAG 132

QY	5896	CTTGAATATTTAAATCTGTAC	5917
Db	131	CTTGAATATTTAAATCTGTAC	110
RESULT	10		
LOCUS	H99724/c		
DEFINITION	H99724	400 bp mRNA linear EST 15-DEC-1995	
ACCESSION	H99724	IMAGE:262972 3', mRNA sequence.	
VERSION	H99724.1	GI:1124392	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 400)		
	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hulman, M., Kucab, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rickin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.		
TITLE	The Mashu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu High quality sequence stops: 360 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert length: 929 Std Error: 0.00 Seq primer: m13 -40 forward High quality sequence stop: 360. Location/Qualifiers 1. 400 /organism="Homo sapiens" /db_xref="GDB:3872614" /db_xref="taxon:9606" /clone="IMAGE:262972" /clone_lib="Soares melanocyte 2NbHM" /sex="Male" /tissue_type="melanocyte" /lab_host="DH10B (ampicillin resistant)" /note="Vector: pRT73D (Pharmacia) with a modified polylinker; Site_1: Not I - Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGGAGCGCGGCAATTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT73 vector (Pharmacia). Library constructed by Bento Soares and M. Regina Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."		
BASE COUNT	139 a	61 c	73 g 126 t 1 others
ORIGIN			

	Query Match	1.2%	Score 82;	DB 10;	Length 400;
	Best Local Similarity	100.0%	Pred. No. 4	3e-23;	
	Matches	82;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
OY	5836	TTTCCAAGCTACCTGTTTACATTGTACACGCGACCACTGGCGCTTTTCATCACAAAG	5895		
Db	191	TTTCCAAGCTACTGTTTACATTGTACACGCGACCACTGGCGCTTTTCATCACAAAG	132		
OY	5896	CTTGAATATTTAAATTCGTGTAC	5917		
Db	131	CTTGAATATTTAAATTCGTGTAC	110		

RESULT 11  
AM591975/c

LOCUS AM591975 408 bp mRNA linear EST 22-MAR-2000  
DEFINITION hit6e6.x1 Soares\_NFL\_T-GBC\_S1 Homo sapiens CDNA clone  
IMAGE:2933986 3', mRNA sequence.  
ACCESSION AM591975  
VERSION AM591975  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 408)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-1@mail.nih.gov](mailto:cgapbs-1@mail.nih.gov)  
This clone is available royalty-free through LIND ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -400P from G1bco  
High quality sequence stop: 356.  
Location/Qualifiers  
1..408

SOURCE

FEATURES

	BASE COUNT	150 a	53 c	77 g	128 t			
ORIGIN								
	Query Match	1.2%; Score 82;	DB 9;	Length 408;				
	Best Local Similarity	100.0%;	Pred.	No. 4.3e-23;				
Matches	82; Conservative	0;	Mismatches	0; Indels	0; Gaps			
OY	5836 TTTCCTCAAGCTACTTGTGGTTACATGTAACAACGCAGCACCGCTTGCCGCCCTTTTCATCACCAG	5895       	Dd	273 TTTCCAAGCATCATTGTTTACATGTAACAACGCAGCACCGCTTGCCGCCCTTTTCATCACCAG	214       			
OY	5896 CTGGAATAATTTAAATCTCTGTAC	5917       	Dd	213 CTTGAATAATTTAAATCTCTGTAC	192			
RESULT_12								
AAd42118								
LOCUS	AAd42118	419 bp	mRNA	linear	EST 02-JUN-1997			
DEFINITION	zwwShoB.f1 Soares.total_fetus.NbZHF8_9w Homo sapiens cdna clone							
VERSION	AAAd42118	IMAGE:774015 5', mRNA sequence.						
KEYWORDS	AAAd42118.1	GI:2153996						
SOURCE	EST.							
ORGANISM	Homo sapiens							
	human.							
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 419)							
REFERENCE								

**AUTHORS**  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,  
Schellenberg, K., Stepien, M., Tan, F., Theising, B., White, Y., Wyllie,  
T., Waterston, R. and Wilson, R.  
WashU-Merck EST Project 1997  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28m13 rev2 ET from Amersham  
High quality sequence stop: 372.

**FEATURES**  
Location/Qualifiers  
1..419  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:774015"  
/clone\_1id="Soares\_total\_fetus\_Nb2HF8\_9w"  
/dev\_stage="8-9 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from pooled 8-9 week  
(total) fetus material with a Not I - oligo(dT) primer [5',  
TGTTACCAATCTGAGAGGAGCGCGCTTATTTTCTTTTCTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

**BASE COUNT**  
129 a 76 c 60 g 154 t

**ORIGIN**  
Query Match 1.2%; Score 82; DB 9; Length 419;  
Best Local Similarity 100.0%; Pred. No. 4.2e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Db**  
5836 TTTCGAAGCTACTGTTTACATTTGACCTGGACACCTTGGCGCTTTATCACAAG 5895  
|||||  
223 TTTCGAAGCTACTGTTTACATTTGACCTGGACACCTTGGCGCTTTATCACAAG 282  
|||||

**Qy**  
5896 CTTGAATATTAAATTCGTAC 5917  
|||||

**Db**  
283 CTTGAATATTAAATTCGTAC 304  
|||||

**RESULT 13**  
AA609977 434 bp mRNA linear EST 02-MAR-1998  
LOCUS AA609977/c  
DEFINITION af09h02.s1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:1031187  
3', mRNA sequence.  
ACCESSION AA609977  
VERSION AA609977.1 GI:2458405  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 434)  
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krimmen, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin,  
J., Moore, B., Schellenberg, K., Stepien, M., Tan, F., Theising, B.,  
White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
WashU-Merck EST Project  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 871 Std Error: 0.00  
Seq primer: -40m13 fwd, ET from Amersham.

**FEATURES**  
Location/Qualifiers  
1..434  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1031187"  
/clone\_1id="Soares\_testis\_NHT"  
/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech Laboratories  
, Inc., and primed with a Not I - oligo(dT) primer [5',  
TGTTACCAATCTGAGAGGAGCGCGCTTATTTTCTTTTCTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

**BASE COUNT**  
159 a 60 c 90 g 125 t

**ORIGIN**  
Query Match 1.2%; Score 82; DB 9; Length 434;  
Best Local Similarity 100.0%; Pred. No. 4.2e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**Db**  
5836 TTTCGAAGCTACTGTTTACATTTGACCTGGACACCTTGGCGCTTTATCACAAG 5895  
|||||  
253 TTTCGAAGCTACTGTTTACATTTGACCTGGACACCTTGGCGCTTTATCACAAG 194  
|||||

**Qy**  
5896 CTTGAATATTAAATTCGTAC 5917  
|||||

**Db**  
193 CTTGAATATTAAATTCGTAC 172  
|||||

**RESULT 14**  
N20387 453 bp mRNA linear EST 18-DEC-1995  
LOCUS N20387/c  
DEFINITION yx37f09.s1 Soares melanocyte 2NbM Homo sapiens cDNA clone  
IMAGE:263945 3', mRNA sequence.  
ACCESSION N20387  
VERSION N20387.1 GI:1125342  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 453)  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman,  
M., Hultman, M., Kucaba, T., Le, N., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston,  
R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
High quality sequence stops: 311  
Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: m13 -40 forward  
High quality sequence stop: 311.

FEATURES  
source  
Location/Qualifiers  
1. .453  
/organism="Homo sapiens"  
/db\_xref="GDB:3873587"  
/db\_xref="taxon:9606"  
/clone="IMAGE:263945"  
/sex="Male"  
/tissue.type="melanocyte"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Vector: pT73D (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo(dT) primer [5',  
TCTTACCAATCTGAGTGGAGCGCGCGAGTTTCTTTTCTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library constructed by Bento Soares and  
M.Fatima Bonaldo. RNA from normal foreskin melanocytes  
(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 163 a 68 c 83 g 139 t  
ORIGIN

Query Match 1.2%; Score 82; DB 10; Length 453;  
Best Local Similarity 100.0%; Pred. No. 4.1e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCCAAGCTACTGTTTACATGTACACGACACCTTGGCGCTTTTCATCACAG 5895  
|||||  
Db 191 TTTCCAAGCTACTGTTTACATGTACACGACACCTTGGCGCTTTTCATCACAG 132  
|||||

QY 5896 CTGCAATATTTAAATCTGTAC 5917  
|||||  
Db 131 CTGCAATATTTAAATCTGTAC 110  
|||||

RESULT 15  
LOCUS AI088996/c 456 bp mRNA linear EST 10-NOV-1998  
DEFINITION  
IMAGE:1634841 3', mRNA sequence.  
ACCESSION AI088996  
VERSION AI088996.1 GI:3428055  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 456)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([infoimage.lnl.gov](http://infoimage.lnl.gov)) for further information.  
Insert Length: 1133 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 449.  
Location/Qualifiers  
1. .456  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1634841"  
/clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

FEATURES  
source

a subtractive hybridization reaction. The driver was  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares NBHSF pool 1:  
309384-310919, 323208-325895 Soares Nb2HP pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares Nb2HP8-9W pool 1:  
758280-760583, 772104-774407 Soares NbHPA pool 1:  
304776-306311, 320136-322823, 326280-326653 Soares NBHOR  
pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 166 a 63 c 92 g 135 t  
ORIGIN

Query Match 1.2%; Score 82; DB 9; Length 456;  
Best Local Similarity 100.0%; Pred. No. 4.1e-23;  
Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5836 TTTCCAAGCTACTGTTTACATGTACACGACACCTTGGCGCTTTTCATCACAG 5895  
|||||  
Db 257 TTTCCAAGCTACTGTTTACATGTACACGACACCTTGGCGCTTTTCATCACAG 198  
|||||

QY 5896 CTGCAATATTTAAATCTGTAC 5917  
|||||  
Db 197 CTGCAATATTTAAATCTGTAC 176  
|||||


Search completed: August 2, 2002, 22:41:57  
Job time: 30172 sec

Mon Aug 5 11:51:36 2002

us-08-973-363-10.oli.rst

---

Page 9



**THIS PAGE BLANK (USPTO)**



NKEENKVKESTQKEKEVKEVNEKSENKESKKIPLLDTPVHTTANSEPVPISE  
ESELHOKTFVCKERMYPKALAKQDREPKGISEBEQLEHTROCLIKIGDHTTECL  
KEYTNPOIKOMRKNLWTFVSKFTEPDARLHLKLYAIKKROSOOHNOONISSVNV  
THVIRNDVDRLEKETTNDHSDSDSYSSDHLISQYHDHHRDROGDAYKRSKRRY  
SATSNGDHRDMDHKODSKRYYSKRSKRRLDDHRSRHRNLEBNLKDSDGSHSRSH  
SDRHRHSDHRSSTBYSHHKSRODRYHSDWQMDHRASGSPRSLDQSPYSPLC  
HRSPEHSHDKSTPEHTWSSRKP\*  
4341..4604  
misc\_feature  
/gene="CHD-2"  
/note="short insert found in longer variant mRNA of CHD-2"

BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN

alignment\_scores:  
Quality: 88.00 Length: 88  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x AF004397 ..

Align seg 1/1 to: AF004397 from: 1 to: 6872

```
1 AspGluIleValSerValIshIseuHisLysLysIleLysThrGluLys 17
|||||
4341 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATTAACAGAAAA 4390
17 sgluasnrgluLysProgluProaspIleGlyIleLysLysGluAlaG 34
|||||
4391 AGAAATGAGAAAAAGCCGAGCCAGATATTGTTAAAGAAAGAAAGCTG 4440
34 luGluLysArgGluThrLysGluLysGluLysArgGluLysArg 50
|||||
4441 AAGAAAAAGAGACAGACAAAAAGAAAGAAATTAACAGCAATTGAAAAAG 4490
51 GluLysLysGluLysGluLysLysLysGluLysGluLysAspAsnLys 67
|||||
4491 GGAAAAAGAAAAAGAGGATTAAGAAAGAAATTAACAGAAAAAGATTAATA 4540
67 sgluLysArgGluLysValLysGluSerThrGlnLysGluLysGluLys 84
|||||
4541 AGAAAAAGAGAAAAACAAAGTAAGAAATCCACACAGAAAGAAAAAGAAAG 4590
84 alLysGluGluLys 88
|||||
4591 TGAAGGAGAGAGAG 4604
```

seq\_name: gb-pat:A58693

seq\_documentation\_block:

LOCUS A58693 265 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 12 from Patent WO9639505.  
ACCESSION A58693  
VERSION A58693.1 GI:3714251  
KEYWORDS  
SOURCE  
ORGANISM  
unclassified.  
REFERENCE 1 (bases 1 to 265)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 12 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES  
Other publication AU 5906996 961224.  
Location/Qualifiers  
1..265  
source  
/organism="unclassified"  
/db\_xref="taxon:32644"

BASE COUNT 158 a 16 c 61 g 30 t  
ORIGIN

alignment\_scores:  
Quality: 75.00 Length: 89  
Ratio: 0.852 Gaps: 1  
Percent Similarity: 98.876 Percent Identity: 98.876

alignment\_block:  
US-08-973-363-11 x A58693 ..

Align seg 1/1 to: A58693 from: 1 to: 265

```
1 AspGluIleValSerValIshIseuHisLysLysIleLysThrGluL 17
|||||
1 GATGAGATTGTTTCAGTGAACATCTACATAAAAAATTAACAGAAAA 50
17 ysgluasnrgluLysProgluProaspIleGlyIleLysLysGluAla 33
|||||
51 AAGAAAAATGAGAAAAAGCCCTGAGCCAGATATTGTTAAAGAAAGAAAGCT 100
34 GluGluLysArgGluThrLysGluLysGluLysArgGluLysArg 50
|||||
101 GAAAGAAAAAGAGACAGACAAAAAGAAAGAAATTAACAGCAATTGAAAAAG 150
50 ggluLysLysGluLysGluLysLysLysGluLysGluLysAspAsnLys 67
|||||
151 GGAGAAAAAGAAAAAGAGCATTAAGAAAGAAATTAACAGAAAAAGATTAATA 200
67 ysgluLysArgGluLysValLysGluSerThrGlnLysGluLysGlu 83
|||||
201 AAGAAAAAGAGAAAAACAAAGTAAGAAATCCACACAGAAAGAAAAAGAA 250
84 ValLysGluGluLys 88
|||||
251 GTGAAGGAGAGAGAG 265
```

seq\_name: gb-pat:A58694

seq\_documentation\_block:

LOCUS A58694 137 bp DNA linear PAT 06-MAR-1998  
DEFINITION Sequence 13 from Patent WO9639505.  
ACCESSION A58694  
VERSION A58694.1 GI:3714252  
KEYWORDS  
SOURCE  
ORGANISM  
unclassified.  
REFERENCE 1 (bases 1 to 137)  
AUTHORS Griffiths,R. and Tiwari,B.  
TITLE AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS  
JOURNAL Patent: WO 9639505-A 13 12-DEC-1996;  
COMMENT ISIS INNOVATION (GB)  
FEATURES  
Other publication AU 5906996 961224.  
Location/Qualifiers  
1..137  
source  
/organism="unclassified"  
/db\_xref="taxon:32644"

BASE COUNT 76 a 10 c 33 g 18 t  
ORIGIN

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x A58694 ..

Align seg 1/1 to: A58694 from: 1 to: 137

```
35 GluLysArgGluThrLysGluLysGluLys 45
|||||
```



104 GAAAAAGACACACAAAAAGAAAAATTAAG 136

seq\_name: gb\_hcg:AC012577

seq\_documentation\_block:

LOCUS AC012577 164144 bp DNA linear HTG 30-MAR-2000

DEFINITION Homo sapiens clone RP11-134, WORKING DRAFT SEQUENCE, 47 unordered

pieces.

ACCESSION AC012577

VERSION AC012577.3 GI:7341804

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 164144)

Britten, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens, clone RP11-134

2 (bases 1 to 164144)

Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barina, N., Beckert, R., Boguslavsky, L., Bouknight, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,

Cooke, P., Deatellano, K., Dewar, K., Domingo, M., Donelan, L., Doyle, M.,

Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,

Galligan, J., Garayana, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,

Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, D.,

Lehoczky, J., Lien, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGuirk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Testa, S., Tirelli, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-OCT-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 30, 2000 this sequence version replaced gi:6479098.

All repeats were identified using RepeatMasker:

Smit, A. P. A. &amp; Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIDR

Web site: http://www.seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2500

Center clone name: L2500

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 136646 bases at least Q40

Consensus quality: 147222 bases at least Q30

Consensus quality: 152549 bases at least Q20

Insert size: 173000; agarose-fp

Insert size: 159544; sum-of-contigs

Quality coverage: 3.1 in Q20 bases; sum-of-contigs

Quality coverage: 3.3 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 47 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1085 1184: contig of 1084 bp in length

\* 1185 2467: contig of 1283 bp in length

\* 2468 2567: gap of 100 bp

\* 2568 4182: contig of 1615 bp in length

\* 4183 4282: gap of 100 bp

\* 4283 5706: contig of 1424 bp in length

\* 5707 5806: gap of 100 bp

\* 5807 7051: contig of 1245 bp in length

\* 7052 7151: gap of 100 bp

\* 7152 8775: contig of 1624 bp in length

\* 8776 8875: gap of 100 bp

\* 8876 10671: contig of 1796 bp in length

\* 10672 10771: gap of 100 bp

\* 10772 12406: contig of 1635 bp in length

\* 12407 12506: gap of 100 bp

\* 12507 13970: contig of 1464 bp in length

\* 13971 14070: gap of 100 bp

\* 14071 15920: contig of 1850 bp in length

\* 15921 16020: gap of 100 bp

\* 16021 18236: contig of 2216 bp in length

\* 18237 18336: gap of 100 bp

\* 18337 19151: contig of 1479 bp in length

\* 19152 21938: contig of 2023 bp in length

\* 21939 22038: gap of 100 bp

\* 22039 23976: contig of 1938 bp in length

\* 23977 24076: gap of 100 bp

\* 24077 25970: contig of 1894 bp in length

\* 25971 26070: gap of 100 bp

\* 26071 27824: contig of 1754 bp in length

\* 27825 27924: gap of 100 bp

\* 27925 30049: contig of 2125 bp in length

\* 30050 30149: gap of 100 bp

\* 30150 31707: contig of 1558 bp in length

\* 31708 31807: gap of 100 bp

\* 31808 33988: contig of 2181 bp in length

\* 33989 34088: gap of 100 bp

\* 34089 36181: contig of 2093 bp in length

\* 36182 36281: gap of 100 bp

\* 36282 39576: contig of 3295 bp in length

\* 39577 39676: gap of 100 bp

\* 39677 42561: contig of 2885 bp in length

\* 42562 42661: gap of 100 bp

\* 42662 45123: contig of 2462 bp in length

\* 45124 45223: gap of 100 bp

\* 45224 48511: contig of 3288 bp in length

\* 48512 48611: gap of 100 bp

\* 48612 51312: contig of 2701 bp in length

\* 51313 51412: gap of 100 bp

\* 51413 54343: contig of 2931 bp in length

\* 54344 54443: gap of 100 bp

\* 54444 57005: contig of 2562 bp in length

\* 57006 57105: gap of 100 bp

\* 57106 60893: contig of 3788 bp in length

\* 60894 60993: gap of 100 bp

\* 60994 63781: contig of 2788 bp in length

\* 63782 63881: gap of 100 bp

\* 63882 66724: contig of 2843 bp in length

\* 66725 66824: gap of 100 bp

\* 66825 69259: contig of 2435 bp in length

\* 69260 69359: gap of 100 bp

\* 69360 72510: contig of 3151 bp in length

\* 72511 72610: gap of 100 bp

\* 72611 76509: contig of 3899 bp in length

\* 76510 76609: gap of 100 bp

\* 76610 80140: contig of 3531 bp in length

\* 80141 80240: gap of 100 bp

\* 80241 84178: contig of 3938 bp in length

\* 84179 84278: gap of 100 bp

\* 84279 89364: contig of 5086 bp in length

\* 89365 89464: gap of 100 bp

\* 89465 94500: contig of 5036 bp in length

\* 94501 94600: gap of 100 bp

\* 94601 99471: contig of 4871 bp in length

\* 99472 99571: gap of 100 bp

\* 99572 104300: contig of 4729 bp in length



```

* as soon as it is available and the accession number will
* be preserved.
*
* 1 1828 contig of 1828 bp in length
* 1829 1928: gap of 100 bp
* 1929 26417: contig of 24489 bp in length
* 26418 26517: gap of 100 bp
* 26518 54437: contig of 27920 bp in length
* 54438 54537: gap of 100 bp
* 54538 101442: contig of 46887 bp in length
* 101425 101524: gap of 100 bp
* 101525 177957: contig of 76453 bp in length.
FEATURES
source
1..177957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1fb="RPC1-11 Human Male BAC"
1..1828
/note="assembly_fragment"
clone_end:T7
vector_side:right"
1929..26417
/note="assembly_fragment"
26518..54437
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
54538..101424
/note="assembly_fragment"
101525..177957
/note="assembly_fragment"
BASF COUNT 53204 a 34244 c 35114 g 54595 t 400 others
ORIGIN
alignment_scores:
Quality: 11.00 Length: 11
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-973-363-11 x AC025298 ..
Align seg 1/1 to: AC025298 from: 1 to: 177957
66 AsnTgSgiuTysArTgIuAsnTysAVallySgiu 76
|||||
53821 AATTAAGAGAAAAAGCAGATTAAGTCACAGAA 53853
seq_name: gb_pr:AC092372
seq_documentation_block:
LOCUS AC092372 101220 bp DNA linear PRI 07-DEC-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.
VERSION AC092372
AC092372.3 GI:17402768
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 101220)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Unpublished
2 (bases 1 to 101220)
DOE Joint Genome Institute.
Direct Submission
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 101220)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
REFERENCE
AUTHORS
REFERENCE
AUTHORS
TITLE

```

```

JOURNAL      Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT      On Dec 7, 2001 this sequence version replaced gi:15290448.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center
              www.shgc.stanford.edu
              Quality: Phrap Quality >=40 100% of Sequence:
              Estimated Total Number of Errors is 0.
              NOTE: This insert is not the entire sequence of the clone (entire
              sequence is 146,7kb). It is clipped at the overlap with AC012624.
              The number of bases overlapped is 90404.
FEATURES
    source
        1..101220
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="RP11-58M12"
BASE COUNT   34122 a 18862 c 17827 g 30409 t
ORIGIN
alignment_scores:
    Quality: 10.00          Length: 10
    Ratio: 1.000           Gaps: 0
Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x AC092372/rev ..

Align seg 1/1 to reverse of: AC092372 from: 1 to: 101220
        67  LysGluLysArgGluAsnLysValLysGlu 76
            ||||||||||||||||||||||||||||
25212  AAGAGAAAAGAGAGAAACAAAGAAAGAA 25183

seq_name: gb_PR:AC012624

seq_documentation_block:
LOCUS      AC012624          134365 bp    DNA          linear    PRI 21-JUL-2001
DEFINITION      Homo sapiens chromosome 5 clone CMD-2082117, complete sequence.
ACCESSION      AC012624
VERSION        AC012624.6   GI:1493679
KEYWORDS       HTG.
SOURCE        human.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 134365)
              DOE Joint Genome Institute and Stanford Human Genome Center.
              Direct Submission
              Unpublished
              2 (bases 1 to 134365)
              DOE Joint Genome Institute.
              Direct Submission
              Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
              3 (bases 1 to 134365)
              DOE Joint Genome Institute and Stanford Human Genome Center.
              Direct Submission
              Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
              4 (bases 1 to 134365)
              DOE Joint Genome Institute and Stanford Human Genome Center.
              Direct Submission
              Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
              On Jul 21, 2001 this sequence version replaced gi:14277267.
COMMENT      Location/Qualifiers
              1..134365
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="5"

```

BASE COUNT 40414 a 24497 c 25503 g 43951 t  
ORIGIN /clone="CTD-2082117"

alignment\_scores: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x AC012624 ..

Align seg 1/1 to: AC012624 from: 1 to: 134365

67 LysgylusArgGluAsnLysValLysGlu 76  
|||||  
119971 AAGAGAAAAGAGAAACAACTAAAGCA 120000

seq\_name: gb\_hlg:AC021449

seq\_documentation\_block:  
LOCUS AC021449 143079 bp DNA linear HTG 10-SEP-2000  
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered  
pieces.  
AC021449  
VERSION AC021449.3 GI:10047806  
KEYWORDS HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Boguslavsky,L., Bouckhalter,B., Brown,A., Burkett,G., Castle,A.,  
Cheong,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,  
DeAratano,K., Dewar,K., Domino,M., Doyle,M., Fenesstor,J.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardina,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lander,T., Lehoczeky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
Macdonald,P., Margulis,N., McKwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meidirim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and zody,M.  
DIRECT SUBMISSION  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:7407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L5154  
Center clone name: 58\_M\_12  
----- Summary Statistics  
Sequencing vector: MJ3; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q30  
Consensus quality: 140814 bases at least Q20  
Insert size: 144000; agarose-fp  
Insert size: 142179; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 38820: contig of 38820 bp in length  
38821 38920: gap of 100 bp  
38921 40411: contig of 1491 bp in length  
40412 40511: gap of 100 bp  
40512 43279: contig of 2768 bp in length  
43280 43379: gap of 100 bp  
43380 46905: contig of 3526 bp in length  
46906 47005: gap of 100 bp  
47006 51830: contig of 4825 bp in length  
51831 51930: gap of 100 bp  
51931 62619: contig of 10689 bp in length  
62620 62719: gap of 100 bp  
62720 75408: contig of 12689 bp in length  
75409 75508: gap of 100 bp  
75509 92516: contig of 17008 bp in length  
92517 92616: gap of 100 bp  
92617 106409: contig of 13793 bp in length  
106410 106509: gap of 100 bp  
106510 143079: contig of 36570 bp in length.  
Location/Qualifiers  
1. 143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-58M12"  
/clone\_1lb="RPCT-11 Human Male BAC"  
1. 38820  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
38921. 40411  
/note="assembly\_fragment"  
40512. 43279  
/note="assembly\_fragment"  
43380. 46905  
/note="assembly\_fragment"  
47006. 51830  
/note="assembly\_fragment"  
51931. 62619  
/note="assembly\_fragment"  
62720. 75408  
/note="assembly\_fragment"  
75509. 92516  
/note="assembly\_fragment"  
92617. 106409  
/note="assembly\_fragment"  
106510. 143079  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

alignment\_scores: 10.00 Length: 10  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-11 x AC021449

Align seg 1/1 to: AC021449 from: 1 to: 143079

67 LysGluysArgGluAsnLysValLysGlu 76  
 |||||  
 117867 AAGAGAAAAAGAGAAACAAAGTAAAGAA 117896

seq\_name: gb\_hlg:AC008531

seq\_documentation\_block:

LOCUS AC008531 145659 bp DNA linear HTG 14-FEB-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
 7 ordered pieces.

ACCESSION AC008531.3 GI:12830078

VERSION HTG; HTGS\_PHASE2; HTGS\_DRAFT.

KEYWORDS human.

ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 145659)

AUTHORS DOE Joint Genome Institute.

TITLE Sequencing of Human Chromosome 5

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 145659)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 On Feb 14, 2001 this sequence version replaced g1:528342.COMMENT -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov-----  
 Project Information  
 Center Project Name: 369535  
 Center clone name: CIT-HSPC\_480B11-----  
 Summary Statistics  
 Consensus quality: 139128 bases at least Q40  
 Consensus quality: 142556 bases at least Q30  
 Consensus quality: 143744 bases at least Q20  
 Estimated insert size: 14800; pulse field gel estimation  
 Estimated insert size: 145059; sum-of-contigs estimation  
 Quality coverage: 6.27 in Q20 bases; pulse field gel estimation  
 Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.NOTE: This is a 'working draft' sequence. It currently  
 consists of 7 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces  
 is believed to be correct as given, however the sizes  
 of the gaps between them are based on estimates that have  
 been provided by the submitter.\* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.\* 1 56174: contig of 56174 bp in length  
 \* 56175 56274: gap of unknown length  
 \* 56275 100874: contig of 44600 bp in length  
 \* 100875 100974: gap of unknown length  
 \* 100975 113127: contig of 12153 bp in length  
 \* 113128 113227: gap of unknown length  
 \* 113228 118190: contig of 4963 bp in length  
 \* 118191 118290: gap of unknown length  
 \* 118291 119694: contig of 1404 bp in length  
 \* 119695 119794: gap of unknown length  
 \* 119795 123297: contig of 3503 bp in length  
 \* 123298 123398: gap of unknown length  
 \* 123399 145659: contig of 22262 bp in length.

\* Location/Qualifiers

FEATURES

## source

1. 145659

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="CTC-480B11"

/clone\_lib="cattech human BAC library C"

BASE COUNT 42561 a 26305 c 27580 g 48609 t 600 others

## ORIGIN

## alignment\_scores:

Quality: 10.00 Length: 10  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-11 x AC008531

Align seg 1/1 to: AC008531 from: 1 to: 145659

67 LysGluysArgGluAsnLysValLysGlu 76  
 |||||  
 48484 AAGAGAAAAAGAGAAACAAAGTAAAGAA 48513

seq\_name: gb\_hlg:AC091946

seq\_documentation\_block:

LOCUS AC091946 193446 bp DNA linear HTG 09-JUN-2001  
 DEFINITION Homo sapiens chromosome 5 clone RP11-36012, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 33 unordered pieces.

ACCESSION AC091946

VERSION AC091946.1 GI:14333882

KEYWORDS HTG; HTGS\_PHASE1.

SOURCE human.

ORGANISM Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 193446)

AUTHORS DOE Joint Genome Institute.

TITLE Sequencing of Human Chromosome 5

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 193446)

AUTHORS DOE Joint Genome Institute.

TITLE Direct Submission

JOURNAL Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 -----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov-----  
 Project Information  
 Center Project Name: 544799  
 Center clone name: RP11-36012-----  
 Summary Statistics  
 Consensus quality: 157767 bases at least Q40  
 Consensus quality: 180259 bases at least Q30  
 Consensus quality: 184175 bases at least Q20  
 Estimated insert size: 204590; agarose-gel estimation  
 Estimated insert size: 190246; sum-of-contigs estimation  
 Quality coverage: 5.38 in Q20 bases; agarose-gel estimation  
 Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.NOTE: This is a 'working draft' sequence. It currently  
 consists of 33 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.\* 1 1358: contig of 1358 bp in length  
 \* 1359 1458: gap of unknown length

```
* 1459 2667: contig of 1209 bp in length
* 2668 2767: gap of unknown length
* 3823 3823: contig of 1056 bp in length
* 3824 3923: gap of unknown length
* 3924 5920: contig of 1997 bp in length
* 5921 6020: gap of unknown length
* 6021 7494: contig of 1474 bp in length
* 7495 7595: gap of unknown length
* 7595 9097: contig of 1502 bp in length
* 9097 9197: gap of unknown length
* 10261 10260: contig of 1064 bp in length
* 10361 12460: gap of unknown length
* 12461 12560: contig of 2100 bp in length
* 12561 14611: gap of unknown length
* 14612 14711: contig of 2051 bp in length
* 14712 16381: gap of unknown length
* 16382 16481: contig of 1670 bp in length
* 16482 17968: gap of unknown length
* 17969 18068: contig of 1487 bp in length
* 18069 20434: gap of unknown length
* 20435 20534: contig of 2366 bp in length
* 20535 23515: gap of unknown length
* 23516 23615: contig of 2981 bp in length
* 23616 27563: gap of unknown length
* 27564 27663: contig of 3948 bp in length
* 27664 30987: gap of unknown length
* 30988 31087: contig of 3324 bp in length
* 31088 36065: gap of unknown length
* 36066 36165: contig of 4978 bp in length
* 36166 40978: gap of unknown length
* 40979 41078: contig of 4813 bp in length
* 41079 45663: gap of unknown length
* 45664 45763: contig of 4585 bp in length
* 45764 51745: gap of unknown length
* 51746 51845: contig of 5982 bp in length
* 51846 57359: gap of unknown length
* 57360 57459: contig of 5514 bp in length
* 57460 67881: gap of unknown length
* 67882 74133: contig of 10422 bp in length
* 74133 74232: gap of unknown length
* 74233 79795: contig of 6151 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 5563 bp in length
* 87512 87611: gap of unknown length
* 87612 92791: contig of 7616 bp in length
* 92792 92891: gap of unknown length
* 92892 102794: contig of 5180 bp in length
* 102795 102894: gap of unknown length
* 102895 110866: contig of 9903 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 7972 bp in length
* 117572 117671: gap of unknown length
* 117672 123738: contig of 6605 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6067 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 6745 bp in length
* 141545 141644: gap of unknown length
* 141645 169109: contig of 10861 bp in length
* 169110 169209: gap of unknown length
* 169210 193446: gap of unknown length
* 193446 24237: contig of 27465 bp in length.
* 24237 24237: gap of unknown length.
```

## FEATURES

SOURCE

```
1..193446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
```

```
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN
/clone="RP11-36012"
/clone.lib="RPC1 human BAC library 11"
/clone="35252 c 37061 g 60360 t 3202 others"
```

```
alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment\_block:

US-08-973-363-11 x AC091946 ..

Align seg 1/1 to: AC091946 from: 1 to: 193446

67 LysGluLysArgGluAsnLysValIysGlu 76

89378 AAGGAAAAAGAGAAAAACCACTAAAGAA 89407

seq\_name: gb\_pr:AC026778

seq\_documentation\_block:

LOCUS AC026778 195433 bp DNA linear PRI 01-JUN-2001

DEFINITION Homo sapiens chromosome 5 clone CTC-428111, complete sequence.

ACCESSION AC026778

VERSION AC026778.4 GI:14277282

KEYWORDS

HTG.

SOURCE

human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 195433)

DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE Direct Submission

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 195433)

DOE Joint Genome Institute.

JOURNAL Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 195433)

DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL Direct Submission

JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell

Drive, Walnut Creek, CA 94598, USA

COMMENT On Jun 1, 2001 this sequence version replaced gi:13677045.

Draft Sequence Produced by DOE Joint Genome Institute

www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center

Quality: Phrap Quality &gt;=40 99.9% of Sequence;

Estimated total Number of Errors is 0.2.

STS Content:

WI-13675 G23101

SHGC-58345 G38487

SHGC-103595 G57841.

Location/Qualifiers

1..195433

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="CTC-428111"

BASE COUNT 62762 a 37302 c 37040 g 58329 t

ORIGIN

alignment\_scores:

Quality: 10.00 Length: 10

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AC026778/rev ..

Align seg 1/1 to reverse of: AC026778 from: 1 to: 195433

67 LysGluLysArgGluAsnLysValIysGlu 76

```
|||||
28018 AAGAGAAAAAGGAAACAAATAAAGAA 27989
seq_name: gb_pr:AC022121

seq_documentation_block:
LOCUS      219258 bp      DNA      linear      PRI 30-AUG-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.
ACCESSION  AC022121
VERSION    AC022121.6  GI:15375145
KEYWORDS   HTG.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 219258)
            DOE Joint Genome Institute and Stanford Human Genome Center.
            TITLE      Unpublished
            JOURNAL    Direct Submission
            AUTHORS    DOE Joint Genome Institute.
            COMMENT    Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            JOURNAL    Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            JOURNAL    Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
            JOURNAL    On Aug 30, 2001 this sequence version replaced gi:15148108.
            COMMENT    Direct Sequence Produced by DOE Joint Genome Institute
            WWW.JGI.DOE.GOV
            Flushing Completed at Stanford Human Genome Center
            WWW.SHGC.STANFORD.EDU
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.4.
            STS Content:
            WI-5811 G04974
            SHGC-58345 G38487
            SHGC-103595 G57841.
FEATURES             Location/Qualifiers
     source           1..219258
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2007H13"
BASE COUNT          71954 a 42062 c 40933 g 64309 t
ORIGIN
alignment_scores:
    Quality: 10.00      Length: 10
    Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-973-363-11 x AC022121/rev ..
Align seg 1/1 to reverse of: AC022121 from: 1 to: 219258
67  LysGLUysARgLUAsnLysValLysGU 76
|||||
87726 AAGAGAAAAAGGAAACAAATAAAGAA 87697
seq_name: gb_htg:AC092382
seq_documentation_block:
LOCUS      276181 bp      DNA      linear      HTG 03-JUL-2001
DEFINITION Homo sapiens chromosome 5 clone RP11-75H1, WORKING DRAFT SEQUENCE,
ACCESSION  AC092382
VERSION    AC092382.1  GI:14589571
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 276181)
            DOE Joint Genome Institute.
            TITLE      Sequencing of Human Chromosome 5
            JOURNAL    Unpublished
            AUTHORS    DOE Joint Genome Institute.
            COMMENT    Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
            JOURNAL    Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            Project Information
            Center Project Name: 435334
            Center clone name: RPCI-11_75H1
            Summary Statistics
            Consensus quality: 229677 bases at least Q40
            Consensus quality: 256163 bases at least Q30
            Consensus quality: 260799 bases at least Q20
            Estimated insert size: 174820; agarose-fp estimation
            Estimated insert size: 271581; sum-of-contigs estimation
            Quality coverage: 8.62 in Q20 bases; agarose-fp estimation
            Quality coverage: 5.55 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 47 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            1
            1061: contig of 1061 bp in length
            1062 1161: gap of unknown length
            1162 2827: contig of 1666 bp in length
            2828 4227: gap of unknown length
            4228 4327: contig of 1300 bp in length
            4328 5618: gap of unknown length
            5619 5718: contig of 1291 bp in length
            5719 6983: gap of unknown length
            6984 7083: contig of 1265 bp in length
            7084 8422: gap of unknown length
            8423 8522: contig of 1339 bp in length
            8523 9771: gap of unknown length
            9772 9871: contig of 1249 bp in length
            9872 11198: gap of unknown length
            11199 11298: contig of 1327 bp in length
            11299 12414: gap of unknown length
            12415 12514: contig of 1116 bp in length
            12515 13806: gap of unknown length
            13807 13906: contig of 1292 bp in length
            13908 15360: gap of unknown length
            15361 15460: contig of 1454 bp in length
            15461 16567: gap of unknown length
            16568 16767: contig of 1207 bp in length
            16768 18390: gap of unknown length
            18391 18490: contig of 1623 bp in length
            18491 18541: gap of unknown length
            18542 19541: contig of 1051 bp in length
            19543 21156: gap of unknown length
            21156 21156: contig of 1515 bp in length
```

```
* 21157 21256: gap of unknown length
* 21257 22448: contig of 1192 bp in length
* 22449 22548: gap of unknown length
* 22549 24499: contig of 1951 bp in length
* 24500 24599: gap of unknown length
* 24600 26384: contig of 1785 bp in length
* 26385 26484: gap of unknown length
* 26485 27883: contig of 1399 bp in length
* 27884 27984: gap of unknown length
* 27985 30386: contig of 2403 bp in length
* 30387 30486: gap of unknown length
* 30487 32480: contig of 1994 bp in length
* 32481 32580: gap of unknown length
* 32581 35441: contig of 2861 bp in length
* 35442 35541: gap of unknown length
* 35542 37800: contig of 2259 bp in length
* 37801 37900: gap of unknown length
* 37901 40587: contig of 2687 bp in length
* 40588 40687: gap of unknown length
* 40688 44159: contig of 3472 bp in length
* 44160 44259: gap of unknown length
* 44260 46336: contig of 2377 bp in length
* 46337 46736: gap of unknown length
* 46737 50082: contig of 3346 bp in length
* 50083 50182: gap of unknown length
* 50183 53988: contig of 3806 bp in length
* 53989 54088: gap of unknown length
* 54089 56592: contig of 2504 bp in length
* 56593 56692: gap of unknown length
* 56693 61352: contig of 4660 bp in length
* 61353 61452: gap of unknown length
* 61453 64254: contig of 2802 bp in length
* 64255 64354: gap of unknown length
* 64355 68825: contig of 4471 bp in length
* 68826 68925: gap of unknown length
* 68926 74393: contig of 5468 bp in length
* 74394 74494: gap of unknown length
* 74494 80561: contig of 6068 bp in length
* 80562 87266: contig of 6965 bp in length
* 87267 87726: gap of unknown length
* 87727 93600: contig of 5874 bp in length
* 93601 93700: gap of unknown length
* 93701 102024: contig of 8324 bp in length
* 102025 102124: gap of unknown length
* 102125 109905: contig of 7781 bp in length
* 109906 110005: gap of unknown length
* 110006 116737: contig of 6732 bp in length
* 116738 116837: gap of unknown length
* 116838 126797: contig of 9960 bp in length
* 126798 126897: gap of unknown length
* 126898 135114: contig of 8217 bp in length
* 135115 135214: gap of unknown length
* 135215 146713: contig of 11499 bp in length
* 146714 146813: gap of unknown length
* 146814 157894: contig of 11081 bp in length
* 157895 157994: gap of unknown length
* 157995 171752: contig of 13758 bp in length
* 171753 171852: gap of unknown length
* 171853 201687: contig of 29835 bp in length
* 201688 201787: gap of unknown length
* 201788 231268: contig of 29481 bp in length
* 231269 231368: gap of unknown length
* 231369 276181: contig of 44813 bp in length.
```

## FEATURES

## source

1..276181

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/clone="RP11-75H1"

/clone\_lib="RPCT human BAC library 11"

86859 a 51769 c 50026 g 82854 t 4673 others

BASE COUNT  
ORIGIN

alignment\_scores:                   Quality:   10.00           Length:   10  
                                  Ratio:   1.000           Gaps:   0  
Percent Similarity: 100.000       Percent Identity: 100.000

## alignment\_block:

US-08-973-363-11 x AC092382/rev ..

Align seg 1/1 to reverse of: AC092382 from: 1 to: 276181

67 LysGluLysArgGluAsnLysValLysGlu 76

|||||  
211974 AAGAGAAAAAGAGAAACAACTAAAGAA 211945

seq\_name: gb\_pl:CNS01BAC

seq\_documentation\_block:

LOCUS                   CNS01BAC                   720 bp   mRNA   linear   PLN 02-SEP-1999

DEFINITION           Botrytis cinerea strain 74 cDNA library under conditions of

nitrogen deprivation.

ACCESSION           AL113980

VERSION           AL113980.1 GI:5828599

KEYWORDS           cDNA library; nitrogen deprivation.

SOURCE           Botryotinia fuckeliana.

ORGANISM           Botryotinia fuckeliana

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;

Helotiales; Sclerotiniaceae; Botryotinia.

REFERENCE           1 (bases 1 to 720)

AUTHORS           Bilton,F., Levis,C., Fortinl,D., Pradier,J.M. and Brygoo,Y.

TITLE           Direct Submission

JOURNAL           Submitted (01-SEP-1999) Phytopathologie, INRA, route de St Cyr,

78026 Versailles, France

2 (bases 1 to 720)

REFERENCE           Genoscope.

AUTHORS           Direct Submission

TITLE           Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage :

CP 5706 91057 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

COMMENT           The cDNA library to be analyzed within the framework of this

project was created using a Botrytis cinerea strain which was grown

under conditions of nitrogen deprivation, which is the normal

situation for B. cinerea during its development on its host plant.

The library was produced in an oriented direction, in the pSII

vector.

FEATURES           Location/Qualifiers

source           1..720

/organism="Botryotinia fuckeliana"

/strain="74"

/db\_xref="taxon:40559"

/note="Genoscope sequence ID : W25F061"

BASE COUNT       105 a   192 c   128 g   295 t

## ORIGIN

## alignment\_scores:

Quality:   9.00           Length:   9

Ratio:   1.000           Gaps:   0

Percent Similarity: 100.000       Percent Identity: 100.000

## alignment\_block:

US-08-973-363-11 x CNS01BAC/rev ..

Align seg 1/1 to reverse of: CNS01BAC from: 1 to: 720

50 ArgGluLysGluLysGluAsnLys 58

|||||  
657 AGAGAGAGAGAGAGAGAGAGATATAA 631

seq\_name: gb\_in:AB009080

seq\_documentation\_block:









```
1 AspGluIleValSerValIysHisLeuHisLysLysLysIleLysThrGlu.L 17
|||||
1 GATGAGATTGTTTCAGTGAACATCTACTATAAAAAATTAACAGAAAA 50
17 ySGluAsnGluGluLysProGluProAspIleGlyIleLysLysGluAla 33
|||||
51 AAGAAAAAGTGAAGAAAAAGCCTGAGCCAGATTGTTATTAAGAGAAAGCT 100
34 GluGluLysArgGluThrLysGluLysGluAsnLysArgGluLeuLysAr 50
|||||
101 GAAGAAAAAGAGACACAAAAAGAGAAAGAAATTAAGGGAATTGAAAAAG 150
50 GGUuLysLysGluLysGluAspLysLysGluLeuLysGluLysAspAsnL 67
|||||
151 GGAGAAAAAGAAAAAGAGATTAAGAAATTAAGAAAAAGTAAATA 200
67 ySGluLysArgGluAsnLysValLysGluSerThrGluLysGluLysGlu 83
|||||
201 AAGAAAAAGAGAAAAACAAGTAAAGATCCACACAGAAAAAGAAAAAGAA 250
84 ValLysGluGluLys 88
|||||
251 GTGAAGGAAGACAAG 265
```

seq\_name: /SIDSl/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:AA142753

seq\_documentation\_block:

ID AA142753 standard; CDNA; 137 BP.

AA142753:

12-MAR-1997 (first entry)

Chicken CHD-W clone CC14 3' motif.

Bird; sex determination; chromodomain-Helicase-DNA binding 1 Avian;

CHD-W; chromodomain-Helicase-DNA binding on the W chromosome; ss.

Gallus sp.

Key Location/Qualifiers

misc\_difference 52

/\*tag= a

/note= "base 52 disrupts the reading frame for

the translated amino acid sequence given

in Fig 7"

W09639505-A1.

12-DEC-1996.

05-JUN-1996; 96WC-GB01341.

06-JUN-1995; 95GB-0011439.

(ISIS-) ISIS INNOVATION LTD.

Griffiths R, Tiwari B;

WPI; 1997-043127/04.

P-PSDB; AAW08145.

Avian chromodomain-helicase-DNA binding genes determine sex in

birds - used for sex determ. and to control sex of progeny

Disclosure; Fig 7; 76pp; English.

A composite (incomplete) nucleotide sequence (AA142753) and putative

translation (AAW08145) sequence are provided of a motif that is found

applied to the 3' end of chicken CHD-W clone CC14 (see also AA142754).

There are no splice donor or acceptor sites within the motif

suggesting it is a final rather than an intermediary product of

CC splicing. The motif is also found as an insert in some CHD-1A clones

CC (see also AA142752).

XX Sequence 137 BP; 75 A; 10 C; 34 G; 18 T; 0 other;

alignment\_scores:

Quality: 11.00 Length: 11

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AA142753 ..

Align seg 1/1 to: AA142753 from: 1 to: 137

35 GluLysArgGluThrLysGluLysGluAsnLys 45

104 GAAAAAGAGACACAAAGAGAAAGAAATTAAG 136

seq\_name: /SIDSl/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AA142753

seq\_documentation\_block:

ID AAC04835 standard; CDNA; 152 BP.

AAC04835:

06-OCT-2000 (first entry)

Human secreted protein 5' EST, SEQ ID NO: 8910.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

gene therapy; chromosome mapping; ss.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-0200610.

26-FEB-1999; 9905-0122487.

(GEST ) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for

obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

diagnostic, forensic, gene therapy and chromosome mapping procedures -

Claim 1; SEQ ID 8910; 71pp + CD-ROM; English.

The present sequence is one of a large number of 5' ESTs derived from

mRNAs encoding secreted proteins. No ORF has yet been conclusively

identified within the present sequence. The 5' ESTs were prepared from

total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

sequences usually correspond mainly to the 3' untranslated region (UTR)

of the mRNA because they are often obtained from oligo-dT primed cDNA

libraries. Such ESTs are not well suited for isolating cDNA sequences

derived from the 5' ends of mRNAs and even in those cases where longer

cDNA sequences have been obtained, the full 5' UTR is rarely included.

5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

in diagnostic, forensic, gene therapy and chromosome mapping procedures.

They are used to obtain upstream regulatory sequences and to design

expression and secretion vectors.

Sequence 152 BP; 37 A; 33 C; 22 G; 60 T; 0 other;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x AAC04835/rev ..

Align seg 1/1 to reverse of: AAC04835 from: 1 to: 152

52 LysArgGluLysGluAspLysLys 59  
|||||  
107 AAAAAGAGAGAGAGAGAGAGAA 84

seq\_name: /STD1/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH70996

seq\_documentation\_block:  
ID: AAH70996 standard: cDNA: 271 BP.

AAH70996:

19-SEP-2001 (first entry)

Human cervical cancer marker nucleic acid 2270.

Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

Homo sapiens.

MO200142467-A2.

14-JUN-2001.

08-DEC-2000; 2000MO-US3312.

08-DEC-1999; 99US-0169681.

21-DEC-1999; 99US-0171350.

14-MAR-2000; 2000US-0189315.

12-MAY-2000; 2000US-0203791.

09-JUN-2000; 2000US-0210600.

21-JUL-2000; 2000US-0220114.

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Schlegel R, Deeds J, Berger A, Zhao X.

WPI: 2001-375006/39.

New isolated nucleic acid for diagnosing and treating cervical cancer

and for assessing and detecting compounds for treating the cancer -

Claim 1; Page 480; 1051pp; English.

The invention relates to novel genes (AAH68727-AAH73383) associated with

cervical cancer with cytostatic activity. The nucleic acids and encoded

polypeptides are useful to assess if a patient is afflicted with

cervical cancer or has a pre-malignant condition; to monitor the

progression of cervical cancer or a premalignant condition in a patient;

and to select and/or assess the efficacy of a compound or therapy for

inhibiting cervical cancer in a patient. The nucleic acids may also be

useful for gene therapy.

Sequence 271 BP; 98 A; 56 C; 75 G; 41 T; 1 other;

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AAH70996 ..

Align seg 1/1 to: AAH70996 from: 1 to: 271

20 GluLysProGluProAspIle 27  
|||||  
128 GAAGAAAGAGAGAGAGAGAGAA 151

seq\_name: /STD1/gcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABR74794

seq\_documentation\_block:

ID: ABR74794 standard: DNA: 287 BP.

ABR74794:

01-FEB-2002 (first entry)

Human foetal liver single exon nucleic acid probe #23099.

Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

Homo sapiens.

MO200157277-A2.

09-AUG-2001.

30-JAN-2001; 2001MO-US00669.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR.

WPI: 2001-483447/52.

Human genome-derived single exon nucleic acid probes useful for

analyzing gene expression in human foetal liver -

Claim 4; SEQ ID NO 23099; 639pp + sequence listing; English.

The invention relates to a single exon nucleic acid probe for

measuring human gene expression in a sample derived from human foetal

liver. The single exon nucleic acid probes may be used for predicting,

measuring and displaying gene expression in samples derived from human

foetal liver. The present sequence is a single exon nucleic acid

probe of the invention.

Note: The sequence data for this patent did not form part of the

printed specification, but was obtained in electronic format directly

from WIPO at ftp.wipo.int/pub/published\_pcl\_sequences.

Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x ABR74794 ..

Align seg 1/1 to: ABR74794 from: 1 to: 287

49 LysArgGluLysGluLysGlu 56

|||||

16 AAAAGGAGAGAGAGAGAGAG 39

```
seq_name: /SIDSI/gcgdata/hold_geneseg/geneseqn-emb1/NA2001A.DAT.ABA39495
seq_documentation_block:
ID ABA39495 standard; DNA: 287 BP.
XX
AC ABA39495;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #17961 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 4; SEQ ID NO 17961; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x ABA39495 ..
Align seg 1/1 to: ABA39495 from: 1 to: 287

49 LysARGlULySLySGlULySGlu 56
|||||
16 AAAAGGAGAGAAAGAGAGAGAG 39
```

```
seq_name: /SIDSI/gcgdata/hold_geneseg/geneseqn-emb1/NA2001A.DAT.AAK23286
seq_documentation_block:
ID AAK23286 standard; DNA: 287 BP.
XX
AC AAK23286;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 23277.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 23277; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x AAK23286 ..
Align seg 1/1 to: AAK23286 from: 1 to: 287

49 LysARGlULySLySGlULySGlu 56
|||||
16 AAAAGGAGAGAAAGAGAGAGAG 39

seq_name: /SIDSI/gcgdata/hold_geneseg/geneseqn-emb1/NA2001A.DAT.AAK49450
seq_documentation_block:
ID AAK49450 standard; DNA: 287 BP.
XX
AC AAK49450;
```

```

DE DT 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 24007.
DE
XX
XX Human: bone marrow expressed exon: gene expression analysis; probe;
KM microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
XX Homo sapiens.
OS
XX
XX MO200157276-A2.
PD
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001MO-US00668.
XX
XX 04-FEB-2000; 2000US-0180312.
XX
XX 26-MAY-2000; 2000US-0207456.
XX
XX 30-JUN-2000; 2000US-0608408.
XX
XX 03-AUG-2000; 2000US-0632366.
XX
XX 21-SEP-2000; 2000US-0234687.
XX
XX 27-SEP-2000; 2000US-0236359.
XX
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SC, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow.
XX
XX
XX Example 4; SEQ ID NO: 24007; 658bp + Sequence Listing: English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
XX
XX Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;
SQ
XX
XX
XX alignment_scores:
XX          Quality:      8.00          Length:      8
XX          Ratio:      1.000          Gaps:      0
XX Percent Similarity: 100.000      Percent Identity: 100.000
XX
XX alignment_block:
XX US-08-973-363-11 x AAK49450      ..
XX
XX Align seg 1/1 to: AAK49450 from: 1 to: 287
XX
XX 49 LysARGGluLysLysGluLysGlu 56
XX |||||||||||||||||||
XX 16 AAGAGGAGAAAGCAAGCAAGAG 39
XX
XX seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.AA126561
XX
XX seq_documentation_block:
XX ID AA126561 standard; DNA; 287 BP.
XX
XX AA126561;
XX
XX 12-OCT-2001 (first entry)
XX
XX Probe #1694 for gene expression analysis in human cervical cell sample.
XX
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KM cervical cancer; ss.
XX
XX

```

OS Homo sapiens.  
 PN WO200157278-A2.  
 PD 09-AUG-2001.  
 PF 30-JAN-2001: 2001WO-0506670.  
 PR 04-FEB-2000: 2000US-0180312.  
 PR 26-MAY-2000: 2000US-0207456.  
 PR 30-JUN-2000: 2000US-0608408.  
 PR 03-AUG-2000: 2000US-0532366.  
 PR 21-SEP-2000: 2000US-0234687.  
 PR 27-SEP-2000: 2000US-0236359.  
 PR 04-OCT-2000: 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn Sq, Hanzel DK, Chen W, Rank DR;  
 DR WPI, 2001-488901/53.  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human cervical epithelial cells -  
 PS Claim 25; SEQ ID No 16494; 487bp; English.  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENP). The present sequence is one such probe. The SENPs are derived  
 CC from human HeLa cells. The SENPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [http://wipo.int/pub/published\\_pcr\\_sequences](http://wipo.int/pub/published_pcr_sequences).  
 SQ Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other.  
 Alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-11 x AAI26561 ..  
 Align seg 1/1 to: AAI26561 from: 1 to: 287  
 49 LysATGAGTulytlysgLulysGlu 56  
 |||||||  
 16 AAAGGAGGAGAGAGAGAGAGAGAG 39  
 seq\_name: /SISL/gsgdata/hold-geneseq/geneseqn-emb1/WA2001A.DAT-AAI55319  
 seq\_documentation\_block:  
 ID-AAI55319 standard; DNA: 287 BP.  
 AC AAI55319;  
 DT 17-OCT-2001 (first entry)  
 XX Probe #24005 used to measure gene expression in human placenta sample.  
 XX Probe: microarray; human; placenta; antenatal diagnosis;  
 XX genetic disorder; ss.  
 XX Homo sapiens.  
 XX WO200157272-A2.  
 XX

```
PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 24005; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 287 BP; 139 A; 5 C; 138 G; 5 T; 0 other;

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x AA155319 ..

Align seg 1/1 to: AA155319 from: 1 to: 287

49 LysATgGluLysLysGluLysGlu 56
   |||||||
16 AAAAGGAGAAAGCAAGAGAGAG 39

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH72066
seq_documentation_block:
ID AAH72066 standard; cDNA; 314 BP.
XX
XX AAH72066;
XX
XX 19-SEP-2001 (first entry)
XX
XX Human cervical cancer marker nucleic acid 3340.
XX
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200142467-A2.
XX
XX 14-JUN-2001.
XX
XX 08-DEC-2000; 2000WO-US33312.
XX
XX 08-DEC-1999; 99US-0169681.
XX 21-DEC-1999; 99US-0171350.
XX 14-MAR-2000; 2000US-0189315.
XX 12-MAY-2000; 2000US-0203791.
XX 09-JUN-2000; 2000US-0210600.
XX 21-JUL-2000; 2000US-0220114.
XX
XX
```

```
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX
XX WPI; 2001-375006/39.
XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer
XX and for assessing and detecting compounds for treating the cancer -
XX
XX Claim 1; Page 649; 1051pp; English.
XX
XX
XX The invention relates to novel genes (AAH68727-AAH73383) associated with
XX cervical cancer with cytostatic activity. The nucleic acids and encoded
XX polypeptides are useful: to assess if a patient is afflicted with
XX cervical cancer or has a pre-malignant condition; to monitor the
XX progression of cervical cancer or a premalignant condition in a patient;
XX and to select and/or assess the efficacy of a compound or therapy for
XX inhibiting cervical cancer in a patient. The nucleic acids may also be
XX useful for gene therapy.
XX
XX
SQ Sequence 314 BP; 124 A; 58 C; 82 G; 50 T; 0 other;

alignment_scores:
    Quality: 8.00      Length: 8
    Ratio: 1.000      Gaps: 0
    Percent Similarity: 100.000    Percent Identity: 100.000

alignment_block:
US-08-973-363-11 x AAH72066 ..

Align seg 1/1 to: AAH72066 from: 1 to: 314

20 GluGluLysProGluProAspIle 27
   |||||||
104 GAAGAAAGCCAGACACAGACATA 127

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAH69191
seq_documentation_block:
ID AAH69191 standard; cDNA; 357 BP.
XX
XX AAH69191;
XX
XX 19-SEP-2001 (first entry)
XX
XX Human cervical cancer marker nucleic acid 465.
XX
XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO200142467-A2.
XX
XX 14-JUN-2001.
XX
XX 08-DEC-2000; 2000WO-US33312.
XX
XX 08-DEC-1999; 99US-0169681.
XX 21-DEC-1999; 99US-0171350.
XX 14-MAR-2000; 2000US-0189315.
XX 12-MAY-2000; 2000US-0203791.
XX 09-JUN-2000; 2000US-0210600.
XX 21-JUL-2000; 2000US-0220114.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Deeds J, Berger A, Zhao X;
XX
XX WPI; 2001-375006/39.
XX
XX New isolated nucleic acid for diagnosing and treating cervical cancer
XX
```



PT and for assessing and detecting compounds for treating the cancer -  
XX  
PS Claim 1: Page 186; 1051pp; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.  
XX  
SQ Sequence 357 BP; 125 A; 73 C; 92 G; 59 T; 8 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-11 x AAH69191 ..  
  
Align seg 1/1 to: AAH69191 from: 1 to: 357  
20 GluGluLysProGluProaspIle 27  
|||||  
105 GAAAGAAAGCCAGAACGACATG 128  
  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA62258  
seq\_documentation\_block:  
ID ABA62258 standard; DNA; 575 BP.  
XX  
AC ABA62258:  
XX  
XX 01-FEB-2002 (first entry)  
XX  
DE Human foetal liver single exon nucleic acid probe #10563.  
XX  
XX Human; foetal liver; gene expression; single exon nucleic acid probe: ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20015727-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001MO-US000669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0609408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human foetal liver -  
XX  
XX Claim 1: SEQ ID NO 10563; 639bp + sequence listing; English.  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at <http://wipo.int/pub/published-pct-sequences>.  
XX  
SQ Sequence 575 BP; 223 A; 93 C; 188 G; 71 T; 0 other;  
  
alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
  
alignment\_block:  
US-08-973-363-11 x ABA62258 ..  
  
Align seg 1/1 to: ABA62258 from: 1 to: 575  
49 LysArgGluLysLysGluLysGlu 56  
|||||  
69 AAAAGCAGACAGAACGACAGAG 92  
  
seq\_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:ABA29604  
seq\_documentation\_block:  
ID ABA29604 standard; DNA; 575 BP.  
XX  
AC ABA29604:  
XX  
XX 23-JAN-2002 (first entry)  
XX  
DE Probe #8070 for gene expression analysis in human heart cell sample.  
XX  
XX Human; gene expression; heart; microarray; vascular system; probe;  
XX cardiovascular disease; hypertension; cardiac arrhythmia;  
XX congenital heart disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PE 30-JAN-2001; 2001MO-US000666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0609408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
XX Claim 1: SEQ ID NO 8070; 530bp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system

CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
 CC congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 575 BP; 223 A; 93 C; 188 G; 71 T; 0 other;

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-11 x ABA29604 ..

Align seg 1/1 to: ABA29604 from: 1 to: 575

49 LysArgGluLysGluLysGlu 56  
 |||||  
 69 AAAAGGAGAGAGAGAGAGAG 92

seq\_name: /STDS1/gcgdata/hold\_geneseq/geneseq-emb1/NA2001A.DAT.AAK10591

seq\_documentation\_block:

ID AAK10591 standard; DNA; 575 BP.

AC AAK10591:

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe SEQ ID NO: 10582.

KW Human: brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer; ss.

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

PS Example 4; SEQ ID NO: 10582; 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX brain. They can be used to measure gene expression in brain cell samples,

XX which may enable the diagnosis and improved treatment of nervous system

XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX epilepsy and cancers. The present sequence is one of the probes of the

XX invention.

SQ Sequence 575 BP; 223 A; 93 C; 188 G; 71 T; 0 other;

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-11 x AAK10591 ..

Align seg 1/1 to: AAK10591 from: 1 to: 575

49 LysArgGluLysGluLysGlu 56  
 |||||  
 69 AAAAGGAGAGAGAGAGAGAG 92

Mon Aug 5 11:51:38 2002

us-08-973-363-11.olip2n.rng

Page 9



**THIS PAGE BLANK (USPTO)**



ACCESSION AU053734  
VERSION AU053734.1 GI:4702216  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum.  
REFERENCE Eukaryota; Mycetozoa; Dictyostelidae; Dictyostellum.  
AUTHORS Mori, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M., Yoshino, R., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.  
TITLE Developmental cDNA in Dictyostelium discoideum  
JOURNAL Unpublished (1998)  
COMMENT Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
Email: d402nuesakura.cc.tsukuba.ac.jp  
PROJECT = Dictyostelium discoideum cDNA project in Japan.  
FEATURES  
source 1..290  
location/Qualifiers  
1..290  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SLJ576"  
/clone\_lib="Dictyostelium discoideum SL (H. Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 158 a 28 c 55 g 47 t 2 others  
ORIGIN  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-11 x AU053734 ..  
Align seg 1/1 to: AU053734 from: 1 to: 290  
seq\_name: gb\_est2:BG611198  
seq\_documentation\_block:  
LOCUS BG611198 413 bp mRNA linear EST 18-APR-2001  
DEFINITION 60261217F1 NIH\_MGC\_60 Homo sapiens cDNA clone IMAGE:4737522 5',  
mRNA sequence.  
ACCESSION BG611198  
VERSION BG611198.1 GI:13662569  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 413)  
JOURNAL NIH-MGC http://mgi.nci.nih.gov/  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM1605 row: j column: 19  
High quality sequence stop: 230.  
Location/Qualifiers

source 1..413  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4737522"  
/clone\_lib="NIH\_MGC\_60"  
/tissue\_type="adencarcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);  
site\_1: SfiI (ggcgctcggc); site\_2: SfiI (ggcgatcggc);  
Double-stranded cDNA was prepared from cell line RNA.  
5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGCCATTTAGCC-3' and 3' adaptor  
sequence: 5'-ATCTAGAGCGCGCGCGCGCATG-GT(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC  
library."  
BASE COUNT 176 a 77 c 104 g 56 t  
ORIGIN  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-11 x BG611198 ..  
Align seg 1/1 to: BG611198 from: 1 to: 413  
seq\_name: gb\_est2:BI508663  
seq\_documentation\_block:  
LOCUS BI508663 447 bp mRNA linear EST 29-AUG-2001  
DEFINITION BB170001B10E08 Bee Brain Normalized/Subtracted Library, BB17 Apis  
mellifera cDNA clone BB170001B10E08 5', mRNA sequence.  
ACCESSION BI508663  
VERSION BI508663.1 GI:15359037  
KEYWORDS EST.  
SOURCE honeybee.  
ORGANISM Apis mellifera  
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
AUTHORS Eukaryota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata  
TITLE 1 (bases 1 to 447)  
JOURNAL Whitfield, C.W., Soares, B., Robertson, H.M., Pardinas, J., Liu, L.,  
Smoller, D. and Robinson, G.E.  
COMMENT An Expressed Sequence Tag Resource for Studies of Brain and  
Behavior in the Honey Bee  
Unpublished (2001)  
Contact: Gene E. Robinson  
Department of Entomology  
University of Illinois  
505 S. Goodwin Ave., Urbana, IL 61801, USA  
Tel: 217 265 0309  
Fax: 217 244 3499  
Email: generobi@life.uiuc.edu  
This research was funded by the University of Illinois Critical  
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation  
Award in Functional Genomics to G.E. Robinson and an NSF  
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.  
PCR Primers  
FORWARD: TAATACGACCTACTATAGG  
REVERSE: ATTACCTCCTACTTAAG  
Insert Length: 447 Std Error: 0.00  
Plate: BB170001B10 row: E column: 08

Seq primer: AGCGATACATTCACACAGCA  
High quality sequence stop: 447.  
Location/Qualifiers

Plate: 120 row: M column: 12  
Seq primer: 77  
Class: BAC ends.  
Location/Qualifiers

1. .447  
/organism="Apis mellifera"  
/strain="mixed strains of European bees, predominantly  
A.m. ligustica"  
/db\_xref="taxon:7460"  
/clone="BBI70001B10E08"  
/clone.lib="Bee Brain Normalized/Subtracted Library, BBI7"  
/sex="female"  
/tissue.type="brain"  
/dev\_stage="adult worker honey bee"  
/lab\_host="DH10B"  
/note="Organ: Brain; Vector: p773-Pac; Site: 1; EcorI;  
Site: 2; NotI; This BBI7 cDNA library was generated by  
subtraction of the BBI6 library with 4000 previously  
sequenced clones. The BBI6 library was contributed by the  
Scares laboratory and it was constructed and normalized  
as described by Bonaldo, M.F., Lennan, G. and Scares,  
M.B. (1996), Genome Research 6(9): 791-806. RNA was  
prepared from dissected brains of adult worker bees of  
various ages and various behavioral groups."

144 a 93 c 85 g 147 t  
BASE COUNT  
ORIGIN  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-11 x AZ257993 ..  
Align seg 1/1 to: AZ257993 from: 1 to: 469

BASE COUNT  
ORIGIN  
alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
alignment\_block:  
US-08-973-363-11 x B1508663/rev ..  
Align seg 1/1 to reverse of: B1508663 from: 1 to: 447

38 gi|tr|lysglulysglulysglu 46  
|||||  
203 GAACAAAGATACACAAAGCAAGCAAGC 229  
seq\_name: gb\_est1:AU053344

75 Lysglulysglulysglulysglu 83  
|||||  
149 AAAGAAAGTACACAAAGCAAGCAAGCA 123

seq\_documentation\_block:  
LOCUS AU053344 510 bp mRNA linear EST 28-APR-1999  
DEFINITION AU053344 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
discoideum cDNA clone SL1432, mRNA sequence.

seq\_name: gb\_gss:AZ257993  
seq\_documentation\_block:  
LOCUS AZ257993 469 bp DNA linear GSS 26-JUL-2000  
DEFINITION RPECI-23-120M12-TV RPECI-23 Mus musculus genomic clone RPECI-23-120M12  
DNA sequence.  
ACCESSION AZ257993  
VERSION AZ257993.1 GI:9463121  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Muskulota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 469)  
Zhao, S., Niernan, W., Feldblum, T., Malek, J., Shatsman, S., Aklnret  
B., Levins, M., McGann, S., Tsagaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C.M.  
Mouse BAC End Sequences from library RPECI-23  
Unpublished (1999)  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPECI-23. For BAC  
library availability, please contact Peter de Jong  
(peterdejong.med.buffalo.edu). Clones may be purchased from  
BACAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)  
or from Resaca Genetics (info@resaca.com). BAC end page:  
http://www.tigr.org/tdb/bac\_ends/mouse/bac\_end\_intro.html

seq\_documentation\_block:  
LOCUS AU053344 510 bp mRNA linear EST 28-APR-1999  
DEFINITION AU053344 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
discoideum cDNA clone SL1432, mRNA sequence.  
ACCESSION AU053344  
VERSION AU053344.1 GI:4701826  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum.  
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.  
1 (bases 1 to 510)  
Mori, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,  
Yoshino, R., Maeda, M., Pl.M., Sato, T., Takemoto, K., Yasukawa, H.,  
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.  
Developmental cDNA in Dictyostelium discoideum  
Unpublished (1998)  
Contact: Hideko Urushihara  
Institute of Biological Sciences  
University of Tsukuba  
3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan  
Email: d402huesakura.cc.tsukuba.ac.jp  
PROJECT - Dictyostelium discoideum cDNA project in Japan.  
Location/Qualifiers  
1. .510  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SL1432"  
/clone.lib="Dictyostelium discoideum SL (H.Urushihara)"  
/dev\_stage="slug"  
BASE COUNT 263 a 64 c 63 g 120 t  
ORIGIN

alignment\_scores:  
 Quality: 9.00 Length: 9  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AU053344 ..

Align seg 1/1 to: AU053344 from: 1 to: 510

31 lysgluAlaGlulysArgGluThr 39

|||||

307 AAGGAAGCTGAGAAAGAGAGAACT 333

seq\_name: gb\_est1:AW790096

seq\_documentation\_block:

LOCUS AW790096 518 bp mRNA linear EST 01-MAY-2001  
 DEFINITION C01520-F lambda Zap, StrataGene Blumeria graminis f. sp. hordei  
 CDNA clone C01520 similar to 101 kd malaria antigen, mRNA sequence.

ACCESSION AW790096 GI:13901693

KEYWORDS EST.

SOURCE Blumeria graminis f. sp. hordei.

ORGANISM Blumeria graminis f. sp. hordei  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;  
 Erysiphales; Erysiphaceae; Blumeria.

REFERENCE 1 (bases 1 to 518)  
 Thomas,S.W., Rasmussen,S.W., Glaring,M.A., Rouster,J.A. and Oliver

AUTHORS R.P.

TITLE Gene identification in the fungal pathogen Blumeria graminis by

JOURNAL expressed sequence tag analysis

COMMENT Unpublished (2000)

Contact: Rasmussen,S.W.  
 Department of Yeast Genetics  
 Carlsberg Laboratory  
 10 Gl. Carlsbergvej, DK-2500, Copenhagen, Denmark

Tel: 45 3327 5230  
 Fax: 45 3327 4766  
 Email: swr@crc.dk

High quality sequence stop: 518

POLYA=No.

FEATURES Location/Qualifiers

1..518  
 /organism="Blumeria graminis f. sp. hordei"

/db\_xref="taxon:62688"

/clone="C01520"

/cell\_type="conidia"

/lab\_host="Hordeum vulgare"

BASE COUNT 131 a 99 c 84 g 204 t

ORIGIN

alignment\_scores:  
 Quality: 9.00 Length: 9  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-08-973-363-11 x AW790096/rev ..

Align seg 1/1 to reverse of: AW790096 from: 1 to: 518

50 ArgGluLysGluLysGluAspLys 58

|||||

451 AAGGAAAGAAAGAAAGAGAGTAAG 425

seq\_name: gb\_gss:AZ879168

seq\_documentation\_block:

LOCUS AZ879168 542 bp DNA linear GSS 05-MAR-2001

DEFINITION RPCI-23-194A5 TJ RPCI-23 Mus musculus genomic clone RPCI-23-194A5,

DATA sequence.  
 accession  
 AZ879168 GI:13197234  
 version  
 AZ879168.1  
 keywords  
 GSS.  
 source  
 house mouse.  
 organism  
 Mus musculus

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhae@tigr.org

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

house mouse.

alignment\_scores:  
 Quality: 9.00 Length: 9  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-11 x AZ879168/rev ..

Align seg 1/1 to reverse of: AZ879168 from: 1 to: 542

41 GluLysGluAsnLysArgGluLeuLys 49

|||||

522 GAGAAAGAGAACAGAGAGAGCTAAAA 496

seq\_name: gb\_est2:BF338481

seq\_documentation\_block:

LOCUS BF338481 791 bp mRNA linear EST 22-NOV-2000

DEFINITION 602034145F2 NCI\_CGAP\_Brn64 Homo sapiens cDNA clone IMAGE:4182227

5', mRNA sequence.

ACCESSION BF338481 GI:11284884

VERSION BF338481.1

KEYWORDS EST.

SOURCE human.



```

ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE         NIH-MGC http://mhc.nci.nih.gov/
JOURNAL       National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT       Unpublished (1999)
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-f@mail.nih.gov
               Tissue Procurement: David N. Louis, M.D.
               CDNA Library Preparation: Life Technologies, Inc.
               DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LHAM9496 row: 1 column: 12
               High quality sequence stop: 304.
               Location/Qualifiers
               1. 791
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="NCI CGAP Brn64"
               /tissue_type="gliblastoma with ESR amplification"
               /lab_host="DH10B (TI phage-resistant)"
               /note="Organ: brain; Vector: PCMV-SPORT6; Site: 1; Note:
               Site: 2; Salt: Cloned unidirectionally; Primer: Oligo dT.
               Average insert size 1.57 kb. Constructed by Life
               Technologies. Note: this is a NCI CGAP Library."

BASE COUNT    412 a 104 c 160 g 115 t
ORIGIN

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-08-973-363-11 x BF338481 ..
  Align seg 1/1 to: BF338481 from: 1 to: 791

52 LysLysGluLysGluAspLysLysGlu 60
|||||
300 AAAAAAGAAAAGACAGCAAGAAAAGAA 326

seq_name: gb_est2:BG533957
seq_documentation_block:
LOCUS      BG533957      817 bp      mRNA      linear      EST 03-APR-2001
DEFINITION 602353058F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4662854 5',
KEYWORDS   mRNA sequence.
ACCESSION  BG533957
VERSION    BG533957.1 GI:13525497
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 817)
AUTHORS   NIH-MGC http://mhc.nci.nih.gov/
TITLE     NIH-MGC http://mhc.nci.nih.gov/
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished (1999)
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-f@mail.nih.gov
               Tissue Procurement: Clontech Laboratories, Inc.
               CDNA Library Preparation: Clontech Laboratories, Inc.
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LHAM9496 row: 1 column: 12
               High quality sequence stop: 304.
               Location/Qualifiers
               1. 791
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="NCI CGAP Brn64"
               /tissue_type="gliblastoma with ESR amplification"
               /lab_host="DH10B (TI phage-resistant)"
               /note="Organ: brain; Vector: PCMV-SPORT6; Site: 1; Note:
               Site: 2; Salt: Cloned unidirectionally; Primer: Oligo dT.
               Average insert size 1.57 kb. Constructed by Life
               Technologies. Note: this is a NCI CGAP Library."

```

```

http://image.llnl.gov
Plate: LHAM9496 row: 1 column: 15
High quality sequence stop: 730.
Location/Qualifiers
1. 817
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI CGAP Brn64"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site: 1;
Site: 2; Salt: Cloned unidirectionally; Primer: Oligo dT.
Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 clones contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH-MGC Library."

BASE COUNT    251 a 139 c 193 g 233 t
ORIGIN

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-08-973-363-11 x BG533957 ..
  Align seg 1/1 to: BG533957 from: 1 to: 817

63 GlnLysAspAsnLysGluLysArgGlu 71
|||||
251 GAAAAGCAATAGACGAGAAAAGAGAA 277

seq_name: gb_est2:BG612310
seq_documentation_block:
LOCUS      BG612310      817 bp      mRNA      linear      EST 18-APR-2001
DEFINITION 602613632F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4739191 5',
KEYWORDS   mRNA sequence.
ACCESSION  BG612310
VERSION    BG612310.1 GI:13663681
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 817)
AUTHORS   NIH-MGC http://mhc.nci.nih.gov/
TITLE     NIH-MGC http://mhc.nci.nih.gov/
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished (1999)
               Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-f@mail.nih.gov
               Tissue Procurement: DCTD/DTP
               CDNA Library Preparation: Clontech Laboratories, Inc.
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LHAM9496 row: 1 column: 12
               High quality sequence stop: 565.
               Location/Qualifiers
               1. 817
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone_lib="NCI CGAP Brn64"
               /tissue_type="adenocarcinoma"

```

/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);  
Site\_1: SfII (ggccgctggcc); Site\_2: SfII (ggccattagcc  
); Double-stranded cDNA was prepared from cell line RNA.  
5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCCGAGCGCGCCGACATG-dt(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC  
Library."

BASE COUNT 304 a 184 c 194 g 135 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x BG612310 ..

Align seg 1/1 to: BG612310 from: 1 to: 817

75 LysGluSerThrGlnLysGluLysGlu 83  
|||||  
657 AAGGAGACACACACAAAAGACAAAGAG 683

seq\_name: gb\_est2:BF161157

seq\_documentation\_block:  
LOCUS BF161157 908 bp mRNA linear EST 30-OCT-2000  
DEFINITION 601769402F1 NCL\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3968561 5',  
mRNA sequence.  
ACCESSION BF161157  
VERSION BF161157.1 GI:11041264  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 908)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapds-remail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM9197 row: d column: 02  
High quality sequence stop: 620.  
Location/Qualifiers  
1. 908  
/organism="Mus musculus"  
/strain="Czech II (feral)"  
/db\_xref="taxon:10090"  
/clone\_image="3968561"  
/clone\_lib="NCL\_CGAP\_Lu29"  
/tissue\_type="spontaneous tumor, metastatic to mammary.  
stem cell origin."  
/lab\_host="DH10B"  
/note="Organ: lung; Vector: PCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI. Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

FEATURES  
source

BASE COUNT 276 a 186 c 180 g 266 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-11 x BF161157/rev ..

Align seg 1/1 to reverse of: BF161157 from: 1 to: 908

39 ThrLysGluLysGluAsnLysArgGlu 47  
|||||  
860 ACAAGGAGAGAGGAAACACAGAGAGAG 834

seq\_name: gb\_est2:BG435095

seq\_documentation\_block:  
LOCUS BG435095 1140 bp mRNA linear EST 14-MAR-2001  
DEFINITION 602507103F1 NIH\_MGC\_79 Homo sapiens cDNA clone IMAGE:4604160 5',  
mRNA sequence.  
ACCESSION BG435095  
VERSION BG435095.1 GI:13341601  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1140)  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapds-remail.nih.gov  
Tissue Procurement: CLONTECH Laboratories, Inc.  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1339 row: f column: 01  
High quality sequence stop: 233.  
Location/Qualifiers  
1. 1140  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="4604160"  
/clone\_lib="NIH MGC 79"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: placenta; Vector: pDNR-LIB (Clontech);  
Site\_1: SfII (ggccgctggcc); Site\_2: SfII (ggccattagcc  
); 5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATTCTAGAGCCGAGCGCGCCGACATG-dt(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC  
Library."

BASE COUNT 520 a 206 c 341 g 73 t  
ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000



---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 3, 2002, 02:12:57 ; Search time 10310.5 seconds

(without alignments)  
537.850 Million cell updates/sec

Title: US-08-973-363-12

Perfect score: 265  
Sequence: 1 GATGAGATTGTTTCACTGAA.....AAGAGTGAGAGAGAGAG 265

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenBank:  
1: gb\_ba.\*  
2: gb\_bkg.\*  
3: gb\_in.\*  
4: gb\_in.\*  
5: gb\_in.\*  
6: gb\_in.\*  
7: gb\_in.\*  
8: gb\_in.\*  
9: gb\_in.\*  
10: gb\_in.\*  
11: gb\_in.\*  
12: gb\_in.\*  
13: gb\_in.\*  
14: gb\_in.\*  
15: gb\_in.\*  
16: gb\_in.\*  
17: gb\_in.\*  
18: gb\_in.\*  
19: gb\_in.\*  
20: gb\_in.\*  
21: gb\_in.\*  
22: gb\_in.\*  
23: gb\_in.\*  
24: gb\_in.\*  
25: gb\_in.\*  
26: gb\_in.\*  
27: gb\_in.\*  
28: gb\_in.\*  
29: gb\_in.\*  
30: gb\_in.\*  
31: gb\_in.\*  
32: gb\_in.\*  
33: gb\_in.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Length DB	ID	Description
------------	-------------	-----------	----	-------------

1	265	100.0	265	A58693
2	218	82.3	6872	AF004397
3	26	9.8	154919	AF007033
4	26	9.8	192079	AC068141
5	26	9.8	237405	AC068142
6	24	9.1	52433	AC101977
7	24	9.1	101220	AC092372
8	24	9.1	134365	AC012624
9	24	9.1	143079	AC021449
10	24	9.1	145659	AC008531
11	24	9.1	154830	AC092736
12	24	9.1	154157	AF003509
13	24	9.1	157625	AC007934
14	24	9.1	193446	AC001946
15	24	9.1	219258	AC022121
16	24	9.1	276181	AC092382
17	23	8.7	137	A58694
18	23	8.7	9654	U67470
19	23	8.7	80374	T8K14
20	23	8.7	105680	AC006266
21	23	8.7	110184	AC020906
22	23	8.7	150874	AC016326
23	23	8.7	166396	AC023957
24	23	8.7	170885	AL356430
25	23	8.7	174135	AL359637
26	23	8.7	193101	AC009561
27	23	8.7	195837	AT091921
28	23	8.7	198316	AC019157
29	23	8.7	203766	AC009592
30	22	8.3	1541	AF180393
31	22	8.3	1585	AF161311
32	22	8.3	12029	AE001432
33	22	8.3	12029	AE001432
34	22	8.3	14012	AE001433
35	22	8.3	83585	AL606475
36	22	8.3	83783	AC016513
37	22	8.3	92633	PEWALAP1_3
38	22	8.3	104695	AC017784
39	22	8.3	110000	PEWALAP1_0
40	22	8.3	114420	AC108194
41	22	8.3	117240	AC104479
42	22	8.3	121287	AC010332
43	22	8.3	130369	AC074090
44	22	8.3	135005	HS860F19
45	22	8.3	141504	AC022295

## ALIGNMENTS

RESULT	1	265 bp	DNA	Linear	PAT 06-MAR-1998
A58693	Sequence 12 from Patent W09639505.				
LOCUS	A58693				
DEFINITION	Sequence 12 from Patent W09639505.				
ACCESSION	A58693				
VERSION	A58693.1	GI:3714251			
KEYWORDS	unidentified.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 265)				
AUTHORS	Griffiths, R. and Tiwari, B.				
TITLE	AVIAN GHD GENES AND THEIR USE IN METHODS FOR SEX IDENTIFICATION IN BIRDS				
JOURNAL	Patent: WO 9639505-A 12 12-DEC-1996;				
COMMENT	ISIS INNOVATION (GB)				
FEATURES	Other publication AU 5906996 961224.				
source	Location/Qualifiers				
	1..265				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
BASE COUNT	158 a	16 c	61 g	30 t	
ORIGIN					

Query Match 100.0%; Score 265; DB 6; Length 265;  
Best Local Similarity 100.0%; Pred. No. 1.3e-106;  
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGAGATTGTTTCAGTGAACATCTACATATAAATAAAGAGAAAAAATGA 60  
|||||  
Db 1 GATGAGATTGTTTCAGTGAACATCTACATATAAATAAAGAGAAAAAATGA 60

QY 61 AGAAACCGTAGCCAGATATGTTATTAAGAGAGCTAAGAAAAAGAGACAAA 120  
|||||  
Db 61 AGAAACCGTAGCCAGATATGTTATTAAGAGAGCTAAGAAAAAGAGACAAA 120

QY 121 AGAGAGAAAAATTAAGAGGATTAAGAGAGAAAAAGAGAGATTAAGAAA 180  
|||||  
Db 121 AGAGAGAAAAATTAAGAGGATTAAGAGAGAAAAAGAGAGATTAAGAAA 180

QY 181 ATTAAGAAAAAGATTAATTAAGAAAAAGAAAAAGATTAAGATCAGACAGAA 240  
|||||  
Db 181 ATTAAGAAAAAGATTAATTAAGAAAAAGAAAAAGATTAAGATCAGACAGAA 240

QY 241 AGAAAAAGAGTGAAGAGAGAG 265  
|||||  
Db 241 AGAAAAAGAGTGAAGAGAGAG 265

RESULT 2  
AF004397 6872 bp mRNA linear VRT 08-OCT-1997  
LOCUS Gallus gallus chemo-helicase-DNA-binding on the Z chromosome  
DEFINITION protein, variant with hydrophilic domain, (CHD-Z) mRNA, complete  
cgs.  
ACCESSION AF004397  
VERSION AF004397.1 GI:2501845  
KEYWORDS  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
REFERENCE 1 (bases 1 to 6872)  
AUTHORS Griffiths, R. and Korn, R.M.  
JOURNAL Gene 197 (1-2), 225-229 (1997)  
MEDLINE 97473516  
REFERENCE 2 (bases 1 to 6872)  
AUTHORS Griffiths, R. and Korn, R.M.  
TITLE Direct Submission  
JOURNAL Submitted (16-MAY-1997) Zoology, Molecular Lab, Glasgow University,  
Glasgow G12 8QQ, UK  
FEATURES  
source  
1. 6872  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
1. 6872  
/gene="CHD-Z"  
228..5654  
/gene="CHD-Z"  
/function="role in chromatin architecture"  
/note="CHD protein with hydrophilic domain"  
/codon\_start=1  
/product="chemo-helicase-DNA-binding on the Z chromosome  
protein"  
/protein\_id="A060282.1"  
/db\_xref="GI:2501845"  
/translation="MNGHSDSESVRNSGSSSRDDSGSAGSGSGSSSGSSDGGSS  
SOGSSDSESGSSGSSGSESDTSREKQVOAKPKPADSEFWKSPSIIAYORSAY  
LKKOQOQKAASDSSGSEEDSSSEPTKKKKHKKDEDMOMSGSGSSTGS  
DSSEADGKASCSSESESDYEPKNKVKSRKPSIRKSGKSTGOKKROLDSEEE  
DDDEDYKGRSROATYVNSYKAEKRTSDILKPCGSDYVQTEDEFTIEKFPD  
SRIGRKATGASTTTTAVADGPNAGFEKELGEIYIILKMGWSHINTWETET  
LKOONVGMKRLNKKKDDKTRMLKNASPEDEVYNNCOELTDLHKOYOVERII

AFHSNQSAAAGYPDYCKMGLPYSECSMEDGALLIAKFPARIDEYFSRMSQSTTPFKD  
CKYLKOPREFVALKQPSYIGGHESELRLDYOLNGLMHLAHSCWCKSGSCITLADMGIG  
KTJOTISFLNLYGPEFLRLVPLSTYFSGREIOTWAPOMNAVYLIGDITSN  
MITHEMHMPOTRKRIKNNILITTYELLKKSPKLGILNMAFTGVDEAHLRLKNDISLY  
RTLIDPESNNRLLITGPIDNLSAKELMSLKHIFIMPEFSSSEDFEEHKGKRGYAS  
LHKLEPFLRLRVKQDVEKSLPAKVEDILHMEKSAIDOKYIKWILTRNKALSKSGS  
STSGFLNIMELKCCMHCLYLRPPDNEFYNNQOALILRSSGLILDLKILRLR  
ERGNRVLFESQVYRMLDILAEYLYKRPDNEFYNNQOALILRSSGLILDLKILRLR  
FLISTRAGGGINLASADTVIVFDSDPNPNDOAOARHARICOKKOVNRYLVATGS  
VEEDILERRAKKVVLDHLYVORMDTGGKTYLHGTGSPSSSTPNKELSAILKFAE  
LKEPBEDEBPOMDIDELKAEIRENPNPGVLYDDELISQFYANFSPNMDIE  
LEPERNSRMNEETIPESQRRRIESEEKOLEETVLPKRNCAKQISFGSGSRSR  
SRYSOGSDSISITERKRPKRGRPRITPRENINQOALIRRSGLILDLKILRLR  
AVARDELVDKSEDLRLGELVHNGCIRKLNKSSGQERAGRLGKVKPTFRISGV  
UNAKLIVISHEELAPLHKSIPDPEERKRYVIPCHTKAHPIDMGKEDDSNLVGI  
YEVYGSWEMIKKMDPLSTOKILPDPPDKPOAKOLOTRADYLLINKDLARKE  
RLAGAGSKRKRTRNNKKNKAKSKIKELKESDSPOSEKSDDEEDDEEDKDEIVSK  
HLHKIKTEREKERPEPDIGIKREKERETKEREKRELKEREKEDKELKED  
NKEKREKRVESVQKEKVEKVEKVEKVEKVEKVEKVEKVEKVEKVEKVEKVEK  
ESELHOKTFESVQKEKVEKVEKVEKVEKVEKVEKVEKVEKVEKVEKVEKVEK  
KEYTNPEQIKOMKRNLMIFVSKFTEFARLKLKYLKAIKROEOOHNDONISSVN  
THYIRNDVRELKETTNDHDSRDSYSDHLLSOYHDHHRDROGDAYKSDSRKRY  
SAFSGNDHDMOHYKODSRYSYDSKRRKLDHRSRHSNLEGNLKDSSGSHDSH  
SDRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHS  
HRSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHSRHS  
misc\_feature  
4341..4604  
/gene="CHD-Z"  
/note="short insert found in longer variant mRNA of CHD-Z"

BASE COUNT 2446 a 1223 c 1520 g 1683 t  
ORIGIN

Query Match 82.3%; Score 218; DB 5; Length 6872;  
Best Local Similarity 100.0%; Pred. No. 4.8e-86;  
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 AAAAAAATGAAAGAAAGCCCTGACCATTTGCTATTAAGAGAGACCTGAGAAA 107  
|||||  
Db 4387 AAAAAAATGAAAGAAAGCCCTGACCATTTGCTATTAAGAGAGACCTGAGAAA 107

QY 108 AAAGAGACAAAAGAGAGAAATTAAGAGAAATTAAGAGAGAAAGAGAAAG 167  
|||||  
Db 4447 AAAGAGACAAAAGAGAGAAATTAAGAGAAATTAAGAGAGAAAGAGAAAG 167

QY 168 AGATTAAGAAAGATTAAGAAAGAAATTAAGAAAGAAAGAAAGAAAGAAAG 227  
|||||  
Db 4507 AGATTAAGAAAGATTAAGAAAGAAATTAAGAAAGAAAGAAAGAAAGAAAG 227

QY 228 AATCCACACAGAAAGAAAGAGTGAAGAGAGAG 265  
|||||  
Db 4567 AATCCACACAGAAAGAAAGAGTGAAGAGAGAG 265

RESULT 3  
AC100733 154919 bp DNA linear HTG 22-NOV-2001  
LOCUS Mus musculus clone RP24-312H21, WORKING DRAFT SEQUENCE, 17  
DEFINITION unordered pieces.  
AC100733  
AC100733.1 GI:17048100  
ACCESSION AC100733  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sclurognathii; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 154919)  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
JOURNAL Mus musculus, clone RP24-312H21  
MEDLINE Unpublished  
REFERENCE 2 (bases 1 to 154919)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barta, N., Bastien, V., Boguski, V., Bourknight, B.,  
Brown, A., Camarata, J., Campobiano, A., Chang, J., Chazaro, B.,

TITLE  
JOURNAL  
COMMENT

Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Farto, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gadyana, S.,  
Gard, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamet, A., Karatas, A., Kells, C., Labouque, K.,  
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,  
Meneus, L., Mihova, T., Mlenga, Y., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, C., Peterson, K., Phukhang, P., Pierre, N., Pollard, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnapack, R.,  
Seman, S., Severy, F., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Tophan, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zatoun, J., Zemdek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Project Information  
Center project name: L17445  
Center clone name: 312\_H-21

Summary Statistics  
Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 149301 bases at least Q40  
Consensus quality: 151798 bases at least Q30  
Consensus quality: 152705 bases at least Q20  
Insert size: 151000; agarose-fp  
sum-of-contigs  
Quality coverage: 7.0 in Q20 bases; agarose-fp  
Quality coverage: 6.8 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
consists of 17 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 23835: contig of 23835 bp in length  
\* 23836 23935: gap of 100 bp  
\* 23936 25171: contig of 1236 bp in length  
\* 25172 25271: gap of 100 bp  
\* 25272 26348: contig of 1077 bp in length  
\* 26349 26448: gap of 100 bp  
\* 26449 29041: contig of 2593 bp in length  
\* 29042 29141: gap of 100 bp  
\* 29142 31763: contig of 2622 bp in length  
\* 31764 31863: gap of 100 bp  
\* 31864 33051: contig of 1188 bp in length  
\* 33052 33151: gap of 100 bp  
\* 33152 35512: contig of 2361 bp in length  
\* 35513 35612: gap of 100 bp  
\* 35613 37576: contig of 1964 bp in length  
\* 37577 37676: gap of 100 bp  
\* 37677 40717: contig of 3041 bp in length  
\* 40718 40817: gap of 100 bp  
\* 40818 43984: contig of 3167 bp in length  
\* 43985 44084: gap of 100 bp

FEATURES  
source  
\* 44085 48175: contig of 4091 bp in length  
\* 48176 48275: gap of 100 bp  
\* 48276 53706: contig of 5431 bp in length  
\* 53707 53806: gap of 100 bp  
\* 53807 59511: contig of 5705 bp in length  
\* 59512 59611: gap of 100 bp  
\* 59612 72539: contig of 12928 bp in length  
\* 72540 72639: gap of 100 bp  
\* 72640 86593: contig of 13954 bp in length  
\* 86594 86693: gap of 100 bp  
\* 86694 111426: contig of 24733 bp in length  
\* 111427 11526: gap of 100 bp  
\* 11527 1154919: contig of 43393 bp in length.  
Location/Qualifiers  
1. 154919  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="RP24-312H21"  
/clone="lib-RP24-312H21 Male Mouse BAC"  
1. 23835  
/note="assembly-fragment"  
clone\_end:SP6  
vector\_side:left  
23936. 25171  
/note="assembly-fragment"  
25272. 26348  
/note="assembly-fragment"  
26449. 29041  
/note="assembly-fragment"  
29142. 31763  
/note="assembly-fragment"  
31864. 33051  
/note="assembly-fragment"  
33152. 35512  
/note="assembly-fragment"  
35613. 37576  
/note="assembly-fragment"  
37677. 40717  
/note="assembly-fragment"  
40818. 43984  
/note="assembly-fragment"  
44085. 48175  
/note="assembly-fragment"  
/note="assembly-fragment"  
48276. 53706  
/note="assembly-fragment"  
53807. 59511  
/note="assembly-fragment"  
59612. 72539  
/note="assembly-fragment"  
72640. 86593  
/note="assembly-fragment"  
86694. 111426  
/note="assembly-fragment"  
111527. 154919  
/note="assembly-fragment"  
clone\_end:77  
vector\_side:right

BASE COUNT 49030 a 29885 c 29043 g 45353 t 1608 others

ORIGIN

Query Match 9.8% Score 26; DB 2; Length 154919;  
Best Local Similarity 100.0%; Pred. No. 0.14;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 AAAAAATTTAAAAACGAAAAACAAA 56  
|||||  
DB 11250 AAAAAATTTAAAAACGAAAAACAAA 11275

RESULT 4  
AC068141 192079 bp DNA linear HTG 29-Apr-2000  
LOCUS AC068141

```
DEFINITION Mus musculus chromosome 5 clone RP23-326E9 strain C57BL6/J, WORKING
ACCESSION AC068141
VERSION AC068141.1 GI:7670125
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 192079)
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
Huang,M.C., Idol,J., Lee-Lin,S.-O., Maduro,Q.L., Maduro,V.B.,
Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
Stantirlop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
Tlonsong,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Wetherby,K.D. and Green,E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 192079)
Green,E.D.
Direct Submission
Submitted (29-APR-2000) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Galtersburg, MD 20877, USA
----- Genome Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
Project Information
Center clone name: 326E09
Center project name: yu
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 187247 bases at least Q40
Consensus quality: 188227 bases at least Q30
Consensus quality: 188721 bases at least Q20
Insert size: 177000; agarose-gel
Insert size: 186000; pulse-field-gel
Insert size: 192079; sum-of-ctrls
Quality coverage: 8.09x in Q20 bases; agarose-gel
Quality coverage: 7.69x in Q20 bases; pulse-field-gel
Quality coverage: 7.45x in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
6760: contig of 6760 bp in length
6761 20444: gap of unknown length
6761 20444: contig of 13684 bp in length
20445 36459: gap of unknown length
20445 36459: contig of 16015 bp in length
36460 58509: gap of unknown length
36460 58509: contig of 22050 bp in length
58510 81567: gap of unknown length
58510 81567: contig of 23058 bp in length
81568 115183: gap of unknown length
81568 115183: contig of 33616 bp in length
115184 149813: gap of unknown length
115184 149813: contig of 34630 bp in length
149814 192079: gap of unknown length
149814 192079: contig of 42266 bp in length.
location/Qualifiers
1. 192079
/organism="Mus musculus"
/strain="C57BL6/J"
```

```
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-326E9"
/clone_1bp="RPCT mouse BAC library 23"
BASE COUNT 58560 a 36452 c 35882 g 61161 t 24 others
ORIGIN
```

```
Query Match 9.8%; Score 26; DB 2; Length 192079;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 32 AAAAAATTAACAGAAAAAGAAAA 57
|||||
Db 121716 AAAAAATTAACAGAAAAAGAAAA 121741
```

```
RESULT 5
AC068142/c 237405 bp DNA linear HTG 29-APR-2000
LOCUS Mus musculus chromosome 5 clone RP23-53F18 strain C57BL6/J, WORKING
DEFINITION DRAFT SEQUENCE, 9 unordered pieces.
ACCESSION AC068142
VERSION AC068142.1 GI:7670126
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 237405)
Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
Huang,M.C., Idol,J., Lee-Lin,S.-O., Maduro,Q.L., Maduro,V.B.,
Mastrian,S.D., McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R.,
Stantirlop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
Tlonsong,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Wetherby,K.D. and Green,E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 237405)
Green,E.D.
Direct Submission
Submitted (29-APR-2000) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Galtersburg, MD 20877, USA
----- Genome Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
Project Information
Center clone name: yt
Center project name: yt
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 230333 bases at least Q40
Consensus quality: 232224 bases at least Q30
Consensus quality: 233239 bases at least Q20
Insert size: 212000; agarose-gel
Insert size: 249000; pulse-field-gel
Insert size: 237405; sum-of-ctrls
Quality coverage: 7.28x in Q20 bases; agarose-gel
Quality coverage: 6.20x in Q20 bases; pulse-field-gel
Quality coverage: 6.50x in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
```



```

* be preserved.
* 1 4718: contig of 4718 bp in length
* 4719 11420: contig of 6702 bp in length
* 11421 28006: contig of 16586 bp in length
* 28007 48211: contig of 20205 bp in length
* 48212 75969: contig of 27758 bp in length
* 75970 110435: contig of 34466 bp in length
* 110436 145778: contig of 35343 bp in length
* 145779 180307: contig of 34529 bp in length
* 180308 237405: contig of 57098 bp in length.
FEATURES
Source
1. 237405
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-53F18"
/clone_11b="RPCI mouse BAC library 23"
BASE COUNT 72271 a 44891 c 45742 g 74492 t 9 others
ORIGIN

```

```

Query Match 9.8%; Score 26; DB 2; Length 237405;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 32 AAAAAATTAACGAAAAAGAAA 57
DB 173891 AAAAAATTAACGAAAAAGAAA 173866

```

```

RESULT 6
AC101977 52433 bp DNA linear HTG 23-NOV-2001
LOCUS Mus musculus clone RP24-333M7, LOW-PASS SEQUENCE SAMPLING.
AC101977
AC101977.1 GI:17060753
VERSION HTG, HTGS_PHASE0.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 52433)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-333M7
Unpublished
2 (bases 1 to 52433)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouknight, B.,
Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B.,
Choe, P., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Deaellano, K., Dewar, K., Diaz, J., Dodge, S., Fero, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gairdner, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagde, B., Harford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamet, A., Karatas, A., Kells, C., Labrecque, K.,
Lamazzara, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marcus, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPherson, R., Melidin, J.,
Menus, L., Mihov, T., Minge, V., Murphy, T., Naylor, J., Nguyen, C.,
North, C., Norman, C., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliyer, J., Peterson, K., Phunhahng, P., Pierre, N., Pollara, V.,
Raymond, C., Retter, R., Rieback, M., Riley, R., Risse, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnapack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

```

# TITLE JOURNAL

Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: 333\_M\_7  
Center clone name: 333\_M\_7

NOTE: This record contains 67 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

```

1 754: contig of 754 bp in length
755 854: gap of 100 bp
855 1486: contig of 632 bp in length
1487 1586: gap of 100 bp
1587 2163: contig of 577 bp in length
2164 2263: gap of 100 bp
2264 2968: contig of 705 bp in length
2969 3068: gap of 100 bp
3069 3774: contig of 706 bp in length
3775 3874: gap of 100 bp
3875 4524: contig of 650 bp in length
4525 4624: gap of 100 bp
4625 5301: contig of 677 bp in length
5302 5401: gap of 100 bp
5402 6084: contig of 683 bp in length
6085 6184: gap of 100 bp
6185 6864: contig of 680 bp in length
6865 6964: gap of 100 bp
6965 7634: contig of 670 bp in length
7635 7734: gap of 100 bp
7735 8437: contig of 703 bp in length
8438 8537: gap of 100 bp
8538 9236: contig of 699 bp in length
9237 9336: gap of 100 bp
9337 9994: contig of 658 bp in length
9995 10094: gap of 100 bp
10095 10794: contig of 700 bp in length
10795 10894: gap of 100 bp
10895 11583: contig of 689 bp in length
11584 11683: gap of 100 bp
11684 12362: contig of 679 bp in length
12363 12462: gap of 100 bp
12463 13145: contig of 683 bp in length
13146 13245: gap of 100 bp
13246 13931: contig of 686 bp in length
13932 14031: gap of 100 bp
14032 141703: contig of 672 bp in length
14704 14803: gap of 100 bp
14804 15467: contig of 664 bp in length
15468 15567: gap of 100 bp
15568 16269: contig of 702 bp in length
16270 16369: gap of 100 bp
16370 17042: contig of 673 bp in length

```

```

* 17043 17142: gap of 100 bp
* 17143 17845: contig of 703 bp in length
* 17846 17945: gap of 100 bp
* 17946 18708: contig of 764 bp in length
* 18710 18809: gap of 100 bp
* 18810 19500: contig of 691 bp in length
* 19501 19600: gap of 100 bp
* 19601 20277: contig of 677 bp in length
* 20278 20377: gap of 100 bp
* 20378 21054: contig of 677 bp in length
* 21055 21154: gap of 100 bp
* 21155 21831: contig of 677 bp in length
* 21832 21931: gap of 100 bp
* 21932 22622: contig of 691 bp in length
* 22623 22722: gap of 100 bp
* 22723 23398: contig of 676 bp in length
* 23399 23498: gap of 100 bp
* 23499 24212: contig of 714 bp in length
* 24213 24312: gap of 100 bp
* 24313 24975: contig of 663 bp in length
* 24976 25075: gap of 100 bp
* 25076 25800: contig of 725 bp in length
* 25801 25900: gap of 100 bp
* 25901 26600: contig of 700 bp in length
* 26601 26700: gap of 100 bp
* 26701 27379: contig of 679 bp in length
* 27380 27479: gap of 100 bp
* 27480 28151: contig of 672 bp in length
* 28152 28251: gap of 100 bp
* 28252 28840: contig of 589 bp in length
* 28841 28940: gap of 100 bp
* 28941 29622: contig of 682 bp in length
* 29623 29722: gap of 100 bp
* 29723 30383: contig of 661 bp in length
* 30384 30483: gap of 100 bp
* 30484 31152: contig of 669 bp in length
* 31153 31252: gap of 100 bp
* 31253 31951: contig of 699 bp in length
* 31952 32051: gap of 100 bp
* 32052 32746: contig of 695 bp in length
* 32747 32846: gap of 100 bp
* 32847 33543: contig of 703 bp in length
* 33550 33649: gap of 100 bp
* 33650 34355: contig of 706 bp in length
* 34356 34455: gap of 100 bp
* 34456 35169: contig of 714 bp in length
* 35170 35269: gap of 100 bp
* 35270 35922: contig of 653 bp in length
* 35923 36022: gap of 100 bp
* 36023 36707: contig of 685 bp in length
* 36708 36807: gap of 100 bp
* 36808 37491: contig of 684 bp in length
* 37492 37591: gap of 100 bp
* 37592 38263: contig of 672 bp in length
* 38264 38363: gap of 100 bp
* 38364 39073: contig of 710 bp in length
* 39074 39173: gap of 100 bp
* 39174 39832: contig of 659 bp in length
* 39833 39932: gap of 100 bp
* 39933 40643: contig of 711 bp in length
* 40644 40743: gap of 100 bp
* 40744 41441: contig of 698 bp in length
* 41442 41541: gap of 100 bp
* 41542 42200: contig of 659 bp in length
* 42201 42300: gap of 100 bp
* 42301 42977: contig of 677 bp in length
* 42978 43077: gap of 100 bp
* 43078 43760: contig of 683 bp in length
* 43761 43860: gap of 100 bp
* 43861 44557: contig of 697 bp in length
* 44558 44657: gap of 100 bp
* 44658 45283: contig of 626 bp in length
* 45284 45383: gap of 100 bp

```

## FEATURES

source

```

* 45384 46089: contig of 706 bp in length
* 46090 46189: gap of 100 bp
* 46190 46860: contig of 671 bp in length
* 46861 46960: gap of 100 bp
* 46961 47688: contig of 728 bp in length
* 47689 47788: gap of 100 bp
* 47789 48467: contig of 679 bp in length
* 48468 48567: gap of 100 bp
* 48568 49254: contig of 687 bp in length
* 49255 49354: gap of 100 bp
* 49355 50021: contig of 667 bp in length
* 50022 50121: gap of 100 bp
* 50122 50825: contig of 704 bp in length
* 50826 50925: gap of 100 bp
* 50926 51628: contig of 703 bp in length
* 51629 51728: gap of 100 bp
* 51729 52433: contig of 705 bp in length.
  location/Qualifiers
    1..52433
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="RP24-333M7"

```

```

Query Match      9.1%; Score 24; DB 2; Length 52433;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 29 ATAAATAATATAACGACAAAAA 52

Db 7863 ATAAATAATATAACGACAAAAA 7886

## RESULT 7

AC092372/c 101220 bp DNA linear PRI 07-DEC-2001  
 LOCUS Homo sapiens chromosome 5 clone RP11-58M12, complete sequence.  
 DEFINITION AC092372

AC092372.3 GI:17402768

HTG.  
 human.  
 ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 101220)

DOE Joint Genome Institute and Stanford Human Genome Center.  
 Direct Submission

Unpublished  
 2 (bases 1 to 101220)

DOE Joint Genome Institute.  
 Direct Submission

Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 101220)  
 DOE Joint Genome Institute and Stanford Human Genome Center.  
 Direct Submission

Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA

On Dec 7, 2001 this sequence version replaced gi:15290448.  
 Draft Sequence Produced by DOE Joint Genome Institute

## COMMENT

www.jgi.doe.gov  
 Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 100% of sequence;  
 Estimated total Number of Errors is 0.  
 NOTE: This insert is not the entire sequence of the clone (entire  
 sequence is 146,7Kb). It is clipped at the overlap with AC012624.  
 The number of bases overlapped is 90404.  
 Location/Qualifiers

## FEATURES

source

```

1..101220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-58M12"

```

```

BASE COUNT      34122 a 18862 c 17827 g 30409 t
ORIGIN
Query Match      9.1%; Score 24; DB 9; Length 101220;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 AAAAGAAAAGAGATACAGAGA 180
DB 25255 AAAAGAAAAGAGATACAGAGA 25232

RESULT 8
AC012624      134365 bp      DNA      linear      PRI 21-JUL-2001
DEFINITION Homo sapiens chromosome 3 clone CTD-2082117, complete sequence.
ACCESSION AC012624
VERSION AC012624.6 GI:1493679
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (01-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 134365)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jul 21, 2001 this sequence version replaced gi:14277267.
FEATURES
source
1.134365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2082117"
BASE COUNT      40414 a 24497 c 25503 g 43951 t
ORIGIN
Query Match      9.1%; Score 24; DB 9; Length 134365;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 AAAAGAAAAGAGATACAGAGA 180
DB 119928 AAAAGAAAAGAGATACAGAGA 119951

RESULT 9
AC021449      143079 bp      DNA      linear      HTG 10-SEP-2000
DEFINITION Homo sapiens clone RP11-58M12, WORKING DRAFT SEQUENCE, 10 unordered
pieces
AC021449
AC021449.3 GI:10047806
KEYWORDS HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

```

# REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 143079)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens, clone RP11-58M12  
Unpublished  
2 (bases 1 to 143079)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Becker,R., Bada,F.,  
Bouguelvsky,L., Bounguelvsky,R., Brown,A., Butte,G., Castle,A.,  
Chen,H., Collins,M., Collins,S., Collymore,A., Cooke,P.,  
Delellano,K., Dekar,K., Domino,N., Doyle,H., Fenebor,J.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,  
Garayna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howard,J.C., Johnson,R., Jones,C., Kam,L., Karalas,A., Klein,J.,  
Lander,T., Lebecky,J., Levine,R., Lien,C., Liu,G., Locke,K.,  
Macdonald,P., Marguis,N., McEwen,P., McCurt,A., McKernan,K.,  
McPherson,R., Weidman,J., Menus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,  
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:7407963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RW/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIMR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

Project information

Center project name: L5154

Center clone name: 58.M.12

Sequencing vector: MJ3, M77815, 100% of reads  
Chemistry: Dye-terminator Big Dye, 100% of reads  
Assembly program: Phrap, version 0.960731  
Consensus quality: 134743 bases at least Q40  
Consensus quality: 139227 bases at least Q30  
Consensus quality: 140814 bases at least Q20  
Insert size: 14400; agarose-fp  
Insert size: 142179; sum-of-coverage  
Quality coverage: 4.6 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently  
consists of 10 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

```

1 38820: contig of 38820 bp in length
38821 38920: gap of 100 bp
38921 40411: contig of 1491 bp in length
40412 40511: gap of 100 bp
40512 43279: contig of 2768 bp in length
43280 43379: gap of 100 bp
43380 46905: contig of 3526 bp in length
46906 47005: gap of 100 bp
47006 51830: contig of 4825 bp in length
51831 51930: gap of 100 bp
51931 51936: contig of 10685 bp in length
51937 62720: gap of 100 bp
62720 75408: contig of 12685 bp in length
75409 75508: gap of 100 bp

```

\* 75509 92516: contig of 17008 bp in length  
\* 92517 92616: gap of 100 bp  
\* 92617 106409: contig of 13793 bp in length  
\* 106410 106509: gap of 100 bp  
\* 106510 143079: contig of 36570 bp in length.  
Location/Qualifiers  
1. 143079  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-58M12"  
/clone\_lib="RPC1-11 Human Male BAC"  
misc-feature  
1. 38820  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left"  
38921. 40411  
/note="assembly\_fragment"  
40512. 43279  
/note="assembly\_fragment"  
43380. 46905  
/note="assembly\_fragment"  
47006. 51830  
/note="assembly\_fragment"  
51931. 62619  
/note="assembly\_fragment"  
62720. 75408  
/note="assembly\_fragment"  
75509. 92516  
/note="assembly\_fragment"  
92617. 106409  
/note="assembly\_fragment"  
106510. 143079  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"  
BASE COUNT 43971 a 26246 c 26678 g 45278 t 906 others  
ORIGIN

Query Match 9.1%; Score 24; DB 2; Length 143079;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 157 AAAAGAAAAAGAGATAGAAAGA 180  
|||||  
Db 117824 AAAAGAAAAAGAGATAGAAAGA 117847

RESULT 10  
AC008531 145659 bp DNA linear HTG 14-FEB-2001  
LOCUS Homo sapiens chromosome 5 clone CTC-480B11, WORKING DRAFT SEQUENCE,  
DEFINITION 7 ordered pieces.  
ACCESSION AC008531  
VERSION AC008531.3 GI:12830078  
KEYWORDS HTG; HTGS\_PHASE2; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 145659)  
TITLE DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 5  
AUTHORS Unpublished  
TITLE 2 (bases 1 to 145659)  
JOURNAL DOE Joint Genome Institute.  
REFERENCE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Feb 14, 2001 this sequence version replaced gi:7528342.  
COMMENT -----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI

Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 369535  
Center clone name: CIT-HSPC\_480B11  
-----  
Summary Statistics  
Consensus quality: 139128 bases at least Q40  
Consensus quality: 142556 bases at least Q30  
Consensus quality: 143744 bases at least Q20  
Estimated insert size: 148000; pulse field gel estimation  
Estimated insert size: 145059; sum-of-contigs estimation  
Quality coverage: 6.27 in Q20 bases; pulse field gel estimation  
Quality coverage: 6.4 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 7 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1  
56174: contig of 56174 bp in length  
56175 56274: gap of unknown length  
56275 100874: contig of 44600 bp in length  
100875 100974: gap of unknown length  
100975 113127: contig of 12153 bp in length  
113128 113227: gap of unknown length  
113228 118190: contig of 4963 bp in length  
118191 118290: gap of unknown length  
118291 119694: contig of 1404 bp in length  
119695 119794: gap of unknown length  
119795 123297: contig of 3503 bp in length  
123298 123398: gap of unknown length  
123398 145659: contig of 22262 bp in length.  
Location/Qualifiers  
1. 145659  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="5"  
/clone="CTC-480B11"  
/clone\_lib="Caltech human BAC library C"  
BASE COUNT 42561 a 26309 c 27580 g 48609 t 600 others  
ORIGIN

Query Match 9.1%; Score 24; DB 2; Length 145659;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 157 AAAAGAAAAAGAGATAGAAAGA 180  
|||||  
Db 48441 AAAAGAAAAAGAGATAGAAAGA 48464

RESULT 11  
AC092736 145830 bp DNA linear HTG 04-OCT-2001  
LOCUS Trypanosoma brucei chromosome IX clone RPC193-28616, \*\*\* SEQUENCING  
DEFINITION IN PROGRESS \*\*\*, 1 ordered pieces.  
ACCESSION AC092736  
VERSION AC092736.5 GI:15920106  
KEYWORDS HTG; HTGS\_PHASE2.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma.  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma.  
1 (bases 1 to 145830)  
REFERENCE El-Sayed, N.M., Ghedini, E., Song, J., Larkin, C., Manless, D., Jones, K.,  
Peterson, J., Hou, L., Zhao, H., Mason, T., Mollis, J., Pat, G., Van  
Aken, S., Ullrich, T., Khalak, H.G., Gerard, C., Leech, V., Dillu, E.,  
Melville, S., White, O., Adams, M.D., Donelson, J.E. and Fraser, C.M.

TITLE Trypanosoma brucei GUTat10.1 RPI193-28616 BAC genomic sequence  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 145830)  
AUTHORS El-Sayed, N.M., Khalak, H. and Adams, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (24-JUL-2001) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
On Oct 4, 2001 this sequence version replaced gi:15789236.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
source 1 145830  
Location/Qualifiers  
1 145830  
/organism="Trypanosoma brucei"  
/isolate="GUTat10.1"  
/db\_xref="taxon:5691"  
/chromosome="IX"  
/clone="RPI193-28616"  
BASE COUNT 35876 a 32464 c 37241 g 40249 t  
ORIGIN  
Query Match 9.1%; Score 24; DB 2; Length 145830;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 172 TAGAAGAAATTAAGAAAGAAAGA 195  
|||||  
Db 139024 TAGAAGAAATTAAGAAAGAAAGA 139001  
RESULT 12  
LOCUS AP003509 154157 bp DNA linear HTG 12-APR-2001  
DEFINITION Oryza sativa chromosome 6 clone P0525F01, \*\*\* SEQUENCING IN  
PROGRESS \*\*\*, in ordered pieces.  
ACCESSION AP003509  
VERSION AP003509.1 GI:13603469  
KEYWORDS HTG: HTGS PHASE2.  
SOURCE Oryza sativa (cultivar: Nipponbare) DNA, clone: P0525F01.  
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;  
Ehmeriales; Oryzoideae; Oryza.  
REFERENCE  
TITLE Sasaki, T., Matsumoto, T. and Yamamoto, K.  
JOURNAL Oryza sativa Nipponbare (GA3) genomic DNA, chromosome 6, PAC  
AUTHORS clone: P0525F01  
TITLE Published Only in Database (2001) In press  
JOURNAL 2 (bases 1 to 134157)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (11-APR-2001) Takuji Sasaki, National Institute of  
Agrobiological Resources, Rice Genome Research Program, Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasaki@affrc.affrc.go.jp, URL: http://rpg.dna.affrc.go.jp/  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and

\* the accession number will be preserved.  
FEATURES  
source 1 154157  
Location/Qualifiers  
1 154157  
/organism="Oryza sativa"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/chromosome="6"  
/clone="P0525F01"  
BASE COUNT 43756 a 32256 c 33371 g 44624 t 150 others  
ORIGIN  
Query Match 9.1%; Score 24; DB 2; Length 154157;  
Best Local Similarity 100.0%; Pred. No. 1.1;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 29 ATTAAGAAATTAAGAAAGAAAGAA 52  
|||||  
Db 35211 ATTAAGAAATTAAGAAAGAAAGAA 35234  
RESULT 13  
LOCUS AC007934 157625 bp DNA linear PRI 21-JAN-2000  
DEFINITION Homo sapiens, clone RP11-29A1, complete sequence.  
ACCESSION AC007934  
VERSION AC007934.7 GI:6721208  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
TITLE Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 157625)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baker, J., Baldwin, J., Barne, N., Beckert, R., Benn, J., Brown, A.,  
Castle, A., Cerny, J., Colangelo, M., Collins, S., Collumore, A.,  
Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,  
Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,  
Funke, R., Gage, D., Galagan, J., Gaidyna, S., Gilbert, D., Grant, G.,  
Hagos, B., Heath, D., Horton, L., Howland, J. C., Jones, C., Kann, L.,  
Karunas, A., Lebecky, J., Lien, C., Locke, K., MacDonald, P.,  
Marquis, N., McKean, P., McKus, A., McKernan, K., McLaughlin, J.,  
Meldrum, J., Molla, M., Morris, M., Morrow, J., Mychaleckyj, J.,  
Naylor, J., Nijhoff, M., O'Connor, T., O'Donnell, P., Pavlid, B.,  
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  
Stange, R., Stojanovic, N., Stojanovic, N., Stone, C., Subramanian, A.,  
Tessier, S., Toriella Miller, L., Vassiliev, H., Vo, A., Wagner, A.,  
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.  
Direct Submission  
Submitted (30-JUN-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 157625)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barne, N., Beckert, R., Beda, F.,  
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,  
Chapel, Y., Colangelo, M., Collins, S., Collumore, A., Cooke, P.,  
Dearellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J.,  
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,  
Gaidyna, S., Grant, G., Hagos, B., Heath, D., Horton, L.,  
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, D.,  
Landers, J., Lebecky, J., Levine, R., Lien, C., Liu, G., Locke, K.,  
Macdonald, P., Marquis, N., McKean, P., McKus, A., McKernan, K.,  
Mcneeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,  
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,  
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,  
Roy, A., Santos, R., Severy, P., Spencer, B., Stange, R., Stojanovic, N.,  
Stojanovic, N., Subramanian, A., Talamas, J., Tessier, S., Theodore, J.,  
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,  
Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (21-JAN-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Jan 20, 2000 this sequence version replaced g1:6532078.  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L601  
Center clone name: 29\_A1

FEATURES  
-----  
source Location/Qualifiers  
1. .157625  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="RPC1-11 Human Male BAC"  
repeat\_region complement(79. .165)  
/rpt\_family="MER41B"  
165. .466  
repeat\_region /rpt\_family="Alusg"  
complement(467. .907)  
/rpt\_family="MER41B"  
repeat\_region complement(933. .1099)  
/rpt\_family="MER41D"  
repeat\_region complement(1100. .1209)  
/rpt\_family="HERVK141HERVK1"  
repeat\_region complement(1215. .1469)  
/rpt\_family="MER66-Internal"  
repeat\_region complement(1820. .2044)  
/rpt\_family="MER66-Internal"  
repeat\_region complement(2048. .2153)  
/rpt\_family="MER49"  
repeat\_region complement(2139. .2426)  
/rpt\_family="MER49"  
repeat\_region complement(2550. .2691)  
/rpt\_family="MER66-Internal"  
repeat\_region complement(2819. .2908)  
/rpt\_family="MER4-Internal"  
repeat\_region complement(3119. .3145)  
/rpt\_family="MER4-Internal"  
repeat\_region 3146. .3453  
/rpt\_family="Alusx"  
repeat\_region complement(3454. .3612)  
/rpt\_family="MER4-Internal"  
repeat\_region complement(3797. .3951)  
/rpt\_family="MER4-Internal"  
repeat\_region complement(4004. .4094)  
/rpt\_family="Alusg"  
repeat\_region complement(4095. .4875)  
/rpt\_family="LTR17"  
repeat\_region complement(4876. .5084)  
/rpt\_family="Alusg"  
repeat\_region complement(5102. .5311)  
/rpt\_family="MER4-Internal"  
repeat\_region complement(5146. .5362)  
/rpt\_family="MER83-Internal"  
repeat\_region 5363. .5669  
/rpt\_family="Aluy"  
complement(5670. .5954)  
/rpt\_family="MER83-Internal"  
repeat\_region 5962. .6259  
/rpt\_family="Alusg"  
repeat\_region 6264. .6564  
/rpt\_family="Alusg"  
repeat\_region 6565. .6598  
/rpt\_family="AT-rich"  
repeat\_region 6599. .6777  
/rpt\_family="L1"

---

repeat\_region /rpt\_family="Alusg/x"  
complement(6782. .6914)  
/rpt\_family="MER83-Internal"  
repeat\_region 6913. .6969  
/rpt\_family="MER51-Internal"  
complement(6970. .7269)  
/rpt\_family="Alusg"  
repeat\_region 7270. .7998  
/rpt\_family="MER51-Internal"  
repeat\_region 7999. .8019  
/rpt\_family="(TG)n"  
repeat\_region 8020. .8587  
/rpt\_family="MER51-Internal"  
repeat\_region 8588. .8886  
/rpt\_family="Alusx"  
repeat\_region 8887. .9231  
/rpt\_family="MER51-Internal"  
repeat\_region 9232. .10683  
/rpt\_family="L1PA6"  
repeat\_region 10691. .11132  
/rpt\_family="L1PA6"  
repeat\_region 11147. .12189  
/rpt\_family="MER51-Internal"  
repeat\_region 12215. .12923  
/rpt\_family="MER57-Internal"  
complement(12932. .13235)  
/rpt\_family="Alusg"  
repeat\_region 13242. .13482  
/rpt\_family="MER51B"  
repeat\_region 13531. .13837  
/rpt\_family="Aluy"  
repeat\_region 13843. .14226  
/rpt\_family="MER51A"  
repeat\_region complement(14280. .14992)  
/rpt\_family="LTR8A"  
repeat\_region 15702. .15750  
/rpt\_family="(T)n"  
repeat\_region complement(15791. .15820)  
/rpt\_family="MIR"  
repeat\_region complement(15943. .16110)  
/rpt\_family="L1PA4"  
repeat\_region 16257. .16285  
/rpt\_family="AT-rich"  
repeat\_region complement(17072. .17391)  
/rpt\_family="Alusc"  
repeat\_region 17758. .17806  
/rpt\_family="CT-rich"  
repeat\_region 17873. .17903  
/rpt\_family="(CA)n"  
repeat\_region 18042. .18265  
/rpt\_family="MER112"  
repeat\_region complement(19327. .19729)  
/rpt\_family="LTR1J"  
repeat\_region complement(19733. .20084)  
/rpt\_family="THEIC"  
repeat\_region 20085. .20375  
/rpt\_family="Aluy"  
complement(20392. .20480)  
/rpt\_family="LTR1J"  
repeat\_region 21388. .21554  
/rpt\_family="MIR"  
repeat\_region 21509. .21564  
/rpt\_family="L2"  
repeat\_region 21942. .21984  
/rpt\_family="AT-rich"  
repeat\_region 22283. .22941  
/rpt\_family="L1"  
repeat\_region 24397. .24591  
/rpt\_family="L1"  
complement(24594. .24971)  
/rpt\_family="LTR1C"  
repeat\_region 24980. .25092  
/rpt\_family="L1"

```
repeat_region 25106..25949
repeat_family="L1M1"
repeat_region 25955..26103
/rpt_family="L1"

Query Match
Best Local Similarity 100.0%; Score 24; DB 9; Length 157625;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 33 AAAAATAAAACGAAAAAGAA 56
|||||
Db 28698 AAAAATAAAACGAAAAAGAA 28721

RESULT 14
AC091946 193446 bp DNA linear HTG 09-JUN-2001
LOCUS Homo sapiens chromosome 5 clone RP11-36012, ** SEQUENCING IN
DEFINITION AC091946
ACCESSION AC091946
VERSION AC091946.1 GI:14333882
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 193446)
AUTHORS DOE Joint Genome Institute.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193446)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 544799
Center clone name: RP11_36012
-----
Summary Statistics
Consensus quality: 157767 bases at least Q40
Consensus quality: 180259 bases at least Q30
Consensus quality: 184175 bases at least Q20
Estimated insert size: 204590; agarose-fp estimation
Estimated insert size: 190246; sum-of-contigs estimation
Quality coverage: 5.79 in Q20 bases; agarose-fp estimation
Quality coverage: 5.79 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1358: contig of 1358 bp in length
* 1358: gap of unknown length
* 1458: gap of unknown length
* 1458: contig of 1209 bp in length
* 2668: gap of unknown length
* 2668: contig of 1056 bp in length
* 3824: gap of unknown length
* 3824: contig of 1997 bp in length
* 5921: gap of unknown length
* 5921: contig of 1474 bp in length
* 6021: gap of unknown length
* 6021: contig of 1502 bp in length
* 7495: gap of unknown length
* 7495: contig of 1502 bp in length
* 9097: gap of unknown length
* 9097: contig of 1064 bp in length
* 9197: gap of unknown length
* 9197: contig of 1064 bp in length
```

```
10261 10360: gap of unknown length
* 10261 12460: contig of 2100 bp in length
* 12461 12560: gap of unknown length
* 12561 14611: contig of 2051 bp in length
* 14612 14711: gap of unknown length
* 14712 16381: contig of 1670 bp in length
* 16382 16481: gap of unknown length
* 16482 17968: contig of 1487 bp in length
* 17969 18068: gap of unknown length
* 18069 20434: contig of 2366 bp in length
* 20435 20534: gap of unknown length
* 20535 23315: contig of 2981 bp in length
* 23316 23515: gap of unknown length
* 23516 27564: gap of 3948 bp in length
* 27564 27664: gap of unknown length
* 27664 30988: contig of 3324 bp in length
* 30988 31088: gap of unknown length
* 31088 36065: contig of 4978 bp in length
* 36066 36166: gap of unknown length
* 36166 40978: contig of 4813 bp in length
* 40979 41078: gap of unknown length
* 41079 45663: contig of 4585 bp in length
* 45664 45763: gap of unknown length
* 45764 51745: contig of 5982 bp in length
* 51746 51845: gap of unknown length
* 51846 57359: contig of 5514 bp in length
* 57360 57459: gap of unknown length
* 57460 67981: contig of 10422 bp in length
* 67982 74132: gap of unknown length
* 74133 74232: gap of 6151 bp in length
* 74233 79795: contig of 5563 bp in length
* 79796 79895: gap of unknown length
* 79896 87511: contig of 7818 bp in length
* 87512 87612: gap of unknown length
* 87613 92891: contig of 5180 bp in length
* 92892 92992: gap of unknown length
* 92993 102894: contig of 9903 bp in length
* 102895 102895: gap of unknown length
* 102896 110866: contig of 7972 bp in length
* 110867 110966: gap of unknown length
* 110967 117571: contig of 6605 bp in length
* 117572 123738: contig of 6067 bp in length
* 123739 123838: gap of unknown length
* 123839 130583: contig of 6745 bp in length
* 130584 130683: gap of unknown length
* 130684 141544: contig of 10861 bp in length
* 141545 141644: gap of unknown length
* 141645 169209: contig of 27465 bp in length
* 169210 169209: gap of unknown length
* 169210 193446: contig of 24237 bp in length.
Location/Qualifiers
1..193446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-36012"
/clone_lib="RP11 human BAC library 11"
BASE COUNT 57571 a 35252 c 37061 g 60360 t 3202 others
ORIGIN
Query Match
Best Local Similarity 100.0%; Score 24; DB 2; Length 193446;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AAAAGAAAAGGATAGAGAAGA 180
|||||
Db 89335 AAAAGAAAAGGATAGAGAAGA 89358

RESULT 15
```

AC022121/c  
 LOCUS AC022121 219258 bp DNA linear PRI 30-AUG-2001  
 DEFINITION Homo sapiens chromosome 5 clone CTD-2007H13, complete sequence.  
 AC022121  
 AC022121.6 GI:15375145  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE  
 AUTHORS 1 (bases 1 to 219258)  
 TITLE DOE Joint Genome Institute and Stanford Human Genome Center.  
 JOURNAL Direct Submission  
 REFERENCE 2 (bases 1 to 219258)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission

REFERENCE  
 AUTHORS Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint  
 TITLE Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 JOURNAL 3 (bases 1 to 219258)  
 REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
 AUTHORS Direct Submission

REFERENCE  
 AUTHORS Submitted (11-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
 TITLE Drive, Walnut Creek, CA 94598, USA  
 JOURNAL 4 (bases 1 to 219258)  
 REFERENCE DOE Joint Genome Institute and Stanford Human Genome Center.  
 AUTHORS Direct Submission

REFERENCE  
 AUTHORS Submitted (30-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell  
 TITLE Drive, Walnut Creek, CA 94598, USA  
 JOURNAL On Aug 30, 2001 this sequence version replaced gi:15148108.  
 COMMENT Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.8% of Sequence;  
 Estimated Total Number of Errors is 0.4.  
 STS Content:  
 WI-5811 G04974  
 WI-13675 G23101  
 SHGC-58345 G38487  
 SHGC-103595 G57841.

FEATURES  
 source Location/Qualifiers  
 1..219258  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2007H13"  
 BASE COUNT 71954 a 42062 c 40933 g 64309 t  
 ORIGIN

Query Match 9.1%; Score 24; DB 9; Length 219258;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 157 AAAAGAAAAAGAGATAAGAAAGA 180  
 ||||||||||||||||||  
 Db 87769 AAAAGAAAAAGAGATAAGAAAGA 87746

Search completed: August 3, 2002, 02:15:42  
 Job time: 42712 sec







```

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225215.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226681.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231245.
PR 08-SEP-2000; 2000US-0231411.
PR 08-SEP-2000; 2000US-0231412.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0233397.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0234400.
PR 14-SEP-2000; 2000US-0234401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 28-SEP-2000; 2000US-0235837.
PR 28-SEP-2000; 2000US-0235837.
PR 28-SEP-2000; 2000US-0236367.
PR 28-SEP-2000; 2000US-0236368.
PR 28-SEP-2000; 2000US-0236369.
PR 28-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244674.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM,
WPI; 2001-541565/60.
Nucleic acids encoding 3224 human nervous system antigen polypeptides,
useful for preventing, diagnosing and/or treating nervous system
cancers and metastases.
Disclosure; SEQ ID NO 13769; 1701pp + Sequence Listing; English.
The invention relates to novel genes (ABA11004-ABA21534) and proteins
(ABA14678-ABA18001) useful for preventing, treating or ameliorating
medical conditions e.g. by protein or gene therapy. The genes are
isolated from a range of human tissues disclosed in the specification.
The nucleic acids, proteins, antibodies and (ant)agonists are useful
in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
and ovarian cancer and other cancers of the adrenal gland, bone, bone

```

CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) Immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 393 BP; 121 A; 58 C; 68 G; 146 T; 0 other;

Query Match  
Best Local Similarity 100.0%; Score 23; DB 22; Length 393;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 AACACGAAAAAGAAATGAAGA 63  
Db 391 AACACGAAAAAGAAATGAAGA 369  
|||||

RESULT 4  
ID AA229250 standard; DNA; 930 BP.  
XX  
AC AA229250:  
XX  
DT 28-FEB-2000 (first entry)  
XX  
DE Rifin 3192 gene.  
XX  
XX Rifin 3192 gene:  
KW mammalian expression plasmid VR1050; DNA vaccine; immunisation;  
KW stage specific protein expression; antimicrobial vaccine;  
KW antimicrobial drug; IAF; immunofluorescent antibody testing; ds.  
XX  
XX Plasmodium falciparum.  
OS  
XX  
XX W09955381-A1.  
PN  
XX  
PD 04-NOV-1999.  
XX  
PF 26-APR-1999; 99WO-US09047.  
XX  
PR 24-APR-1998; 98US-0082947.  
XX  
PR 23-APR-1999; 99US-0082947.  
XX  
PA (USNA ) US SEC OF NAVY.  
XX  
PI Hoffman SL, Carucci DJ;  
XX  
DR WPI; 2000-086380/07.  
XX  
XX  
XX Use of microbial, animal and/or human genomic data for identification  
PT of vaccine, drug or diagnostic products -  
XX  
XX  
XX Claim 12; Page 16; 23pp; English.  
XX  
XX The present sequence is the Rifin 3192 gene of chromosome 2. Rifin genes  
CC were used to characterise the protein expression from various life cycle  
CC stages of P. falciparum. Oligonucleotide primers were used to amplify each  
CC selected ORF from chromosome 2 using genomic DNA as template. The  
CC purified amplified products were cloned into the mammalian expression  
CC plasmid VR1050 to produce DNA vaccines for immunisation. Blood and sera  
CC obtained from groups of mice immunised with doses of vaccine was used to  
CC identify stage specific expression of Rifin protein by immunofluorescent  
CC antibody testing. The method is useful for determination of subcellular  
CC localisation of proteins and for the development of antimicrobial  
CC vaccines and drugs.  
XX

SQ Sequence 930 BP; 342 A; 139 C; 178 G; 271 T; 0 other;

Query Match  
Best Local Similarity 100.0%; Score 22; DB 21; Length 930;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 AGAAAAAGAAATGAAGAAA 66  
Db 877 agaaaaagaataatgaagaaaa 898  
|||||

RESULT 5  
ID AA229252  
XX  
AC AA229252:  
XX  
DT 28-FEB-2000 (first entry)  
XX  
DE Rifin 3194 gene.  
XX  
XX Rifin 3194 gene:  
KW mammalian expression plasmid VR1050; DNA vaccine; immunisation;  
KW stage specific protein expression; antimicrobial vaccine;  
KW antimicrobial drug; IAF; immunofluorescent antibody testing; ds.  
XX  
XX Plasmodium falciparum.  
OS  
XX  
XX W09955381-A1.  
PN  
XX  
PD 04-NOV-1999.  
XX  
PF 26-APR-1999; 99WO-US09047.  
XX  
PR 24-APR-1998; 98US-0082947.  
XX  
PR 23-APR-1999; 99US-0082947.  
XX  
PA (USNA ) US SEC OF NAVY.  
XX  
PI Hoffman SL, Carucci DJ;  
XX  
DR WPI; 2000-086380/07.  
XX  
XX  
XX Use of microbial, animal and/or human genomic data for identification  
PT of vaccine, drug or diagnostic products -  
XX  
XX  
XX Claim 14; Page 16; 23pp; English.  
XX  
XX The present sequence is the Rifin 3194 gene of chromosome 2. Rifin genes  
CC were used to characterise the protein expression from various life cycle  
CC stages of P. falciparum. Oligonucleotide primers were used to amplify each  
CC selected ORF from chromosome 2 using genomic DNA as template. The  
CC purified amplified products were cloned into the mammalian expression  
CC plasmid VR1050 to produce DNA vaccines for immunisation. Blood and sera  
CC obtained from groups of mice immunised with doses of vaccine was used to  
CC identify stage specific expression of Rifin protein by immunofluorescent  
CC antibody testing. The method is useful for determination of subcellular  
CC localisation of proteins and for the development of antimicrobial  
CC vaccines and drugs.  
XX  
XX  
XX Sequence 969 BP; 355 A; 144 C; 206 G; 264 T; 0 other;

Query Match  
Best Local Similarity 100.0%; Score 22; DB 21; Length 969;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 45 AGAAAAAGAAATGAAGAAA 66  
Db 916 agaaaaagaataatgaagaaaa 937  
|||||

```

RESULT 6
AAZ29254 ID AAZ29254 standard; DNA; 978 BP.
XX
XX
AC AAZ29254;
XX
XX 28-FEB-2000 (first entry)
XX
XX Rffin 3198 gene.
XX
XX Rffin 3198 gene: P.falciparum chromosome 2; Rffin protein;
XX mammalian expression plasmid VR1050; DNA vaccine; immunisation;
XX stage specific protein expression; antimicrobial vaccine;
XX antimicrobial drug; IAF; immunofluorescent antibody testing; ds.
XX
XX Plasmodium falciparum.
XX
XX WO955381-A1.
XX
XX 04-NOV-1999.
XX
XX 26-APR-1999; 99WO-US09047.
XX
XX 24-APR-1998; 98US-0082947.
XX
XX 23-APR-1999; 99US-0082947.
XX
XX (USNA ) US SEC OF NAVY.
XX
XX Hoffman SL, Carucci DJ;
XX
XX WPI; 2000-086380/07.
XX
XX Use of microbial, animal and/or human genomic data for identification
XX of vaccine, drug or diagnostic products -
XX
XX Claim 16; Page 17; 23pp; English.
XX
XX The present sequence is the Rffin 3198 gene of chromosome 2. Rffin genes
XX were used to characterise the protein expression from various life cycle
XX stages of P.falciparum. Oligonucleotide primers were used to amplify each
XX selected ORF from chromosome 2 using genomic DNA as template. The
XX purified amplified products were cloned into the mammalian expression
XX plasmid VR1050 to produce DNA vaccines for immunisation. Blood and sera
XX obtained from groups of mice immunised with doses of vaccine was used to
XX identify stage specific expression of Rffin protein by immunofluorescent
XX antibody testing. The method is useful for determination of subcellular
XX localisation of proteins and for the development of antimicrobial
XX vaccines and drugs.
XX
XX Sequence 978 BP; 357 A; 152 C; 208 G; 261 T; 0 other;
XX
XX Query Match 8.3%; Score 22; DB 21; Length 978;
XX Best Local Similarity 100.0%; Pred. No. 4.1;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX OY 45 AGAAAAAGAAATGAAGAAAA 66
XX |
XX Db 925 agaaaaaagaatgaagaataa 946
XX
XX RESULT 7
XX AAA70286
XX ID AAA70286 standard; DNA; 984 BP.
XX
XX AAA70286;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:419.
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX

```

```

XX
XX OS Plasmodium falciparum.
XX
XX WO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF) HOFFMAN S.
XX (CARU) CARUCCI D.
XX (GARD) GARDNER M.
XX (VENT) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 575; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially when they are rifins or secreted or membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasite lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
XX
XX Sequence 984 BP; 360 A; 145 C; 183 G; 296 T; 0 other;
XX
XX Query Match 8.3%; Score 22; DB 21; Length 984;
XX Best Local Similarity 100.0%; Pred. No. 4.1;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX OY 45 AGAAAAAGAAATGAAGAAAA 66
XX |
XX Db 931 agaaaaaagaatgaagaataa 952
XX
XX RESULT 8
XX AAA70284
XX ID AAA70284 standard; DNA; 1038 BP.
XX
XX AAA70284;
XX
XX 07-NOV-2000 (first entry)
XX
XX Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:417.
XX
XX Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX

```

```
OS Plasmodium falciparum.
XX
XX WO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 574-575; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially in the detection of infection with P. falciparum, membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasitic lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
XX
XX Sequence 1038 BP; 379 A; 157 C; 209 G; 293 T; 0 other:
XX
XX
XX Query Match 8.3%; Score 22; DB 21; Length 1038;
XX Best Local Similarity 100.0%; Pred. No. 4.1;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 45 AGAAAAAGAAATGAAAGAAA 66
XX ||||||||||||||||||||
XX Db 985 agaaaaaagaatgaagaaaa 1006
XX
XX
XX RESULT 9
XX AAA70279
XX ID AAA70279 standard; DNA; 1047 BP.
XX
XX AC AAA70279;
XX
XX DT 07-NOV-2000 (first entry)
XX
XX DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:412.
XX
XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
XX antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
XX OS Plasmodium falciparum.
XX
```

```
XX
XX WO200025728-A2.
XX
XX 11-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26796.
XX
XX 05-NOV-1998; 98US-0107131.
XX
XX (HOFF/) HOFFMAN S.
XX (CARU/) CARUCCI D.
XX (GARD/) GARDNER M.
XX (VENT/) VENTER J C.
XX
XX Hoffman S, Carucci D, Gardner M, Venter JC;
XX WPI; 2000-365347/31.
XX
XX Proteins encoded by chromosome 2 of the human malarial parasite,
XX Plasmodium falciparum, useful as antimalarial vaccines and in the
XX diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 573; 577pp; English.
XX
XX The present invention describes proteins and their fragments (I) encoded
XX by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
XX Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
XX vaccines against P. falciparum infection comprising (I) or (II).
XX (I) and (II) are useful for the development of vaccines against
XX P. falciparum infection. (I) and polyclonal antisera or a monoclonal
XX antibody raised to immunogens comprising the sequences of (I), are
XX useful in the detection of infection with P. falciparum. Furthermore,
XX (I) (especially in the detection of infection with P. falciparum, membrane proteins)
XX can aid the identification of drugs to treat or prevent P. falciparum
XX infection, or they can be used to identify drug resistance in
XX P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
XX subsequent identification of proteins encoded by it will help to expand
XX our understanding of parasite biology, a process hampered by the
XX complexity of the parasitic lifecycle, and provide new targets for
XX vaccine and drug development. Parasite resistance to drugs and mosquito
XX resistance to insecticides have led to a resurgence of malaria in many
XX parts of the world, and there is a pressing need for vaccines and new
XX drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
XX and protein sequences given in the present invention, but which are not
XX specifically mentioned within the specification.
XX
XX Sequence 1047 BP; 383 A; 163 C; 211 G; 290 T; 0 other:
XX
XX
XX Query Match 8.3%; Score 22; DB 21; Length 1047;
XX Best Local Similarity 100.0%; Pred. No. 4.1;
XX Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 45 AGAAAAAGAAATGACAGAAA 66
XX ||||||||||||||||||||
XX Db 994 agaaaaaagaatcgaagaaaa 1015
XX
XX
XX RESULT 10
XX ABL18816
XX ID ABL18816 standard; DNA; 29783 BP.
XX
XX AC ABL18816;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 7921.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX OS Drosophila melanogaster.
XX
```

PN WO200171042-A2.  
XX  
XX 27-SEP-2001.  
PD  
XX 23-MAR-2001; 2001WO-US09231.  
PF  
XX 23-MAR-2000; 2000US-191637P.  
PR  
XX 11-JUL-2000; 2000US-0614150.  
FR  
XX (PEKE ) PE CORP NY.  
PA  
XX  
XX Venter JC, Adams M, Li PWD, Myers EM;  
PI WPI; 2001-656860/75.  
DR  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PR interactions -  
PT  
XX  
XX Claim 1; SEQ ID NO 7921; 21np + Sequence Listing; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
CC sequences (AB101840-AB116175) and the encoded proteins  
CC (AB57737-AB572072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 29783 BP; 7503 A; 6790 C; 6798 G; 8692 T; 0 other;

Query Match 8.3%; Score 22; DB 23; Length 29783;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 AAAAAATTAACGCAAAAAA 52  
Db 19930 aaaaaataaaacagaaaaa 19951  
|||||

RESULT 11  
AA181813/C  
ID AA181813 standard; cDNA; 471 BP.  
XX  
XX AA181813;  
AC  
XX  
XX 06-NOV-2001 (first entry)  
DT  
XX  
XX Human polynucleotide SEQ ID NO 1873.  
DE  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200164835-A2.  
PN  
XX  
XX 07-SEP-2001.  
PD  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
PF  
XX  
XX 28-FEB-2000; 2000US-0515126.  
PR  
XX 18-MAY-2000; 2000US-0577409.  
XX  
XX (HYSF-) HYSEQ INC.  
PA  
XX

PI Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI; 2001-514838/56.  
DR  
XX P-PSDB; AA001882.  
DR  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PR disorders -  
PT  
XX  
XX Claim 1; SEQ ID NO 1873; 1399pp + Sequence Listing; English.  
PS  
XX  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 471 BP; 121 A; 123 C; 126 G; 97 T; 4 other;

Query Match 7.9%; Score 21; DB 22; Length 471;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 146 AAAAGCGAGAAAAAAGAAAAA 166  
Db 73 AAAAGCGAGAAAAAAGAAAAA 53  
|||||

RESULT 12  
AA144389  
ID AA144389 standard; DNA; 588 BP.  
XX  
XX AA144389;  
AC  
XX  
XX 17-OCT-2001 (first entry)  
DT  
XX  
XX Probe #13075 used to measure gene expression in human placenta sample.  
DE  
XX  
XX Probe; microarray; human; placenta; antenatal diagnosis.  
KW genetic disorder; ss.  
KW  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200157272-A2.  
PN  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 30-JAN-2001; 2001WO-US00663.  
PF  
XX  
XX 04-FEB-2000; 2000US-0180312.  
PR  
XX 26-MAY-2000; 2000US-0207456.  
PR  
XX 30-JUN-2000; 2000US-0608408.  
PR  
XX 03-AUG-2000; 2000US-0632366.  
PR  
XX 21-SEP-2000; 2000US-0234687.  
PR  
XX 27-SEP-2000; 2000US-0236359.  
PR  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-488897/53.  
XX  
XX

```

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
PS
PS Claim 25; SEQ ID No 13075; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENPs).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 588 BP; 280 A; 84 C; 115 G; 109 T; 0 other;
XX
Query Match 7.98; Score 21; DB 22; Length 588;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 172 TAAGAAAGATTAAAGAAA 192
Db 218 taagaagaattaaaagaataa 238
XX
RESULT 13
AAZ29256
ID AAZ29256 standard; DNA; 936 BP.
XX
AC AAZ29256;
XX
DT 28-FEB-2000 (first entry)
XX
DE Rifin 3201 gene.
XX
KW Rifin 3201 gene; P.falciparum chromosome 2; Rifin protein;
KW mammalian expression plasmid VR1050; DNA vaccine; immunisation;
KW stage specific protein expression; antimicrobial vaccine;
KW antimicrobial drug; IAF; immunofluorescent antibody testing; ds.
XX
OS Plasmodium falciparum.
XX
PN WO955381-A1.
XX
PD 04-NOV-1999.
XX
PF 26-APR-1999; 99WO-US09047.
XX
PR 24-APR-1998; 98US-0082947.
XX
PR 23-APR-1999; 99US-0082947.
XX
PA (USNA ) US SEC OF NAVY.
XX
PI Hoffman SL, Carucci DJ;
XX
DR WPI; 2000-086380/07.
XX
PT Use of microbial, animal and/or human genomic data for identification
PT of vaccine, drug or diagnostic products -
PS
PS Claim 18; Page 17; 23pp; English.
XX
CC The present sequence is the Rifin 3201 gene of chromosome 2. Rifin genes
CC were used to characterize the protein expression from various life cycle
CC stages of P.falciparum. Oligonucleotide primers were used to amplify each
CC selected ORF from chromosome 2 using genomic DNA as template. The
CC purified amplified products were cloned into the mammalian expression
CC plasmid VR1050 to produce DNA vaccines for immunisation. Blood and sera
CC obtained from groups of mice immunised with doses of vaccine was used to
CC identify stage specific expression of Rifin protein by immunofluorescent
CC antibody testing. The method is useful for determination of subcellular
CC localisation of proteins and for the development of antimicrobial
CC vaccines and drugs.
XX
SQ Sequence 936 BP; 357 A; 136 C; 190 G; 253 T; 0 other;

```

Query Match	Similarity	100.0%	Pred. No. 11	Length 936
Best Local	Similarity	100.0% <td>Pred. No. 11</td> <td>Length 936</td>	Pred. No. 11	Length 936
Matches 21	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0
OY	46 GAAAAAAGAAATGAAGAAAA	66		
Db	884 gaaaaaagaatgaagaaaa	904		
RESULT 14				
AAH33151				
ID	AAH33151 standard; cDNA; 976 BP.			
XX	AAH33151;			
XX				
DT	03-SEP-2001 (first entry)			
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:207.			
XX				
KN	Human; colon cancer; colon cancer antigen; diagnosis; detection;			
KN	colorectal carcinoma; chromosome 16; ss.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200122920-A2.			
PD	05-APR-2001.			
XX				
PF	28-SEP-2000; 2000WO-US26524.			
PR	29-SEP-1999; 99US-0157137.			
PR	03-NOV-1999; 99US-0163280.			
XX				
PA	(HUMA-) HUMAN GENOME SCI INC.			
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;			
XX				
DR	WPI; 2001-235357/24.			
DR	P-PSDB; AAG73720.			
XX				
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,			
PR	useful for preventing, diagnosing and/or treating colorectal cancers -			
XX				
PS	Claim 1; Page 2352; 9803pp; English.			
XX				
CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon			
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where			
CC	the proteins are collectively known as colon cancer antigens. The colon			
CC	cancer antigens have cytoprotective activity and can be used in gene			
CC	therapy and vaccine production. N and P may be used in the prevention,			
CC	diagnosis and treatment of diseases associated with inappropriate P			
CC	expression. For example, N and P may be used to treat disorders			
CC	associated with decreased expression by rectifying mutations or deletions			
CC	in a patient's genome that affect the activity of P by expressing			
CC	inactive proteins or to supplement the patient's own production of P.			
CC	Additionally, N may be used to produce the colon cancer-associated P,			
CC	by inserting the nucleic acids into a host cell and culturing the cell			
CC	to express the proteins. N and P can be used in the prevention, diagnosis			
CC	and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204			
CC	and AAH77789 represent sequences used in the exemplification of the			
CC	present invention.			
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were			
CC	missing at time of publication, meaning no sequences are present for			
CC	SEQ ID NO:1027 to 1052, 7921 and 7922.			
XX				
SO	Sequence 976 BP; 431 A; 164 C; 204 G; 168 T; 9 other;			
Query Match	7.9%; Score 21; DB 22; Length 976;			
Best Local Similarity	100.0%; Pred. No. 11;			
Matches 21; Conservative	0; Mismatches	0; Indels	0; Gaps	0





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: August 2, 2002, 22:41:57 ; Search time 7016.61 Seconds

(without alignments)  
509,747 Million cell updates/sec

Title: US-08-973-363-12

Perfect score: 265

Sequence: 1 GATGAGATGCTTCACTGCA.....AAGAGTGAAGAGAGAG 265

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 674847542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthm: \*  
3: em\_estin: \*  
4: em\_estov: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hic: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hic: \*  
12: gb\_gss: \*  
13: em\_gss\_hum: \*  
14: em\_gss\_hiv: \*  
15: em\_gss\_pln: \*  
16: em\_gss\_vrt: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	10.2	882	12	AZ184391 SP.1002.B
2	26	9.8	590	12	AZ091555 RPT-23-2
3	24	9.1	343	9	AA748563
4	24	9.1	375	12	AZ213564
5	24	9.1	430	9	A1242163
6	24	9.1	456	9	AA699918
7	24	9.1	485	12	AO642393
8	24	9.1	538	12	AO500417
9	24	9.1	643	12	AO52151
10	24	9.1	657	12	AO508184
11	24	9.1	677	10	BM015387
12	24	9.1	821	10	BF239967
13	23	8.7	372	9	AU039357
14	23	8.7	393	9	AA658426
15	23	8.7	683	9	AL653205
16	23	8.7	695	10	B1640191
17	23	8.7	743	12	AG032665

C 18	23	8.7	826	12	AO912123	AO912123 nbe00171
C 19	23	8.7	867	12	AZ686023	AZ686023 ENT0807F
C 20	23	8.7	920	12	AZ674814	AZ674814 ENT0807F
C 21	22	8.3	154	12	AG059567	AG059567 Pan t10g1
C 22	22	8.3	301	9	A1479119	A1479119 tm31e03.x
C 23	22	8.3	413	12	A2161843	A2161843 SP.0070.LB
C 24	22	8.3	415	10	BM274338	BM274338 PEST00a4
C 25	22	8.3	518	12	TA169E06P	TA169E06P T. brucei
C 26	22	8.3	520	10	BM273317	BM273317 PEST00a4
C 27	22	8.3	554	10	BM275208	BM275208 PEST00a6
C 28	22	8.3	560	10	BM274989	BM274989 PEST00a7
C 29	22	8.3	584	10	BM273877	BM273877 PEST00a6
C 30	22	8.3	616	12	AA550018	AA550018 1097m3.gm
C 31	22	8.3	619	12	AZ831967	AZ831967 2M0112G05
C 32	22	8.3	876	12	AZ539395	AZ539395 EMTF009TF
C 33	22	8.3	884	9	AU067771	AU067771 AU067771
C 34	22	8.3	938	12	CNS01SN3	AL165432 Tetradodon
C 35	22	8.3	950	10	BE961679	BE961679 601647954
C 36	22	8.3	980	12	CNS0450L	AL275646 Tetradodon
C 37	22	8.3	1062	12	CNS02MNV	AL205168 Tetradodon
C 38	22	8.3	1248	10	BC851122	BC851122 1024030D0
C 39	21	7.9	148	10	BR075197	BR075197 224037.WA
C 40	21	7.9	165	9	AM368791	AM368791 IL2.VT019
C 41	21	7.9	208	9	AT632370	AT632370 WB06908.x
C 42	21	7.9	254	9	AA355716	AA355716 EST64158
C 43	21	7.9	273	12	AG059338	AG059338 Pan t10g1
C 44	21	7.9	310	9	AT033270	AT033270 ow/4a06.s
C 45	21	7.9	355	9	AA305433	AA305433 EST176411

## ALIGNMENTS

RESULT 1  
AZ184391 882 bp DNA linear GSS 30-AUG-2000  
SP.1002.B2.G05.F7A Strongylocentrotus purpuratus, purple sea urchin  
LOCUS  
DEFINITION  
clone genomic BAC library Strongylocentrotus purpuratus genomic  
clone Plate-1002 Col-10 Row-N, DNA sequence.

ACCESSION  
AZ184391  
KEYWORDS  
GSS.  
SOURCE  
Strongylocentrotus purpuratus.  
ORGANISM  
Strongylocentrotus purpuratus.  
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
Echinozoa; Echinozoa; Echinodermata; Echinodermata; Echinodermata;  
Strongylocentrotidae; Strongylocentrotus.

REFERENCE  
AUTHORS  
Cameron, R.A., Mahalir, G., Rast, J.P., Martinez, P., Blondi, T.R.,  
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray,  
G.A., Ettensohn, C.A., Lehnach, H., Britten, R.J., Davidson, E.H. and  
Hood, L.  
A sea urchin genome project: Sequence scan, virtual map, and  
additional resources  
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Cameron, RA, Davidson, EH, Hood, L  
Division of Biology 156-29  
California Institute of Technology  
Pasadena California 91225, USA  
Tel: (626) 395-8421  
Fax: (626) 793-3047  
Email: scameron@caltech.edu  
Plate: 1002 row: N column: 10  
Seq primer: 77  
Class: BAC ends  
High quality sequence stop: 882.  
Location/Qualifiers

FEATURES  
source  
1..882  
/organism="Strongylocentrotus purpuratus"  
/db\_xref="taxon:7668"  
/clone="Plate-1002 Col-10 Row-N"  
/clone\_lib="Strongylocentrotus purpuratus, purple sea

urchin, sperm genomic BAC library"					
/note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli-					
DH10B"					
BASE COUNT	377 a	51 c	304 g	67 t	83 others
ORIGIN					

Query Match	10.28;	Score 27;	DB 12;	Length 882;
Best Local Similarity	100.0%;	Pred. No. 3.7;		
Matches 27; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	145	GAAAAGGAGAAAAACAAAAAGAGA	171
Db	532	GAAAAGGAGAAAAAGAAAAAGAGA	558

RESULT	2
AZ091555	
LOCUS	AZ091555
DEFINITION	RPC1-23-27F16.TV RPC1-23 Mus musculus genomic clone RPC1-23-27F16,
ACCESSION	AZ091555
VERSION	AZ091555
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus

REFERENCE  
1 (bases 1 to 590)  
Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akınret  
AUTHORS

TITLE	Mouse BAC End Sequences from Library RPCI-23
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSS: RPCI-23-27F16.TJ

Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: [szhao@tigr.org](mailto:szhao@tigr.org)  
Clones are derived from the mouse BAC library RPCL-23. For BAC library availability, please contact Pieter de Jong ([pieter@dejong.med.buffalo.edu](mailto:pieter@dejong.med.buffalo.edu)). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.html>) or from Resea ch Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end page: [http://www.tigr.org/tbdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tbdb/bac_ends/mouse/bac_end_intro.html)  
plate: 27 row: F column: 16  
Seq primer: T7  
Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .590

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCT-23-27F16"
/clone_id="RPCT_23"
/sex="Female"
/lab_host="DH10B"

```

/note=organ: kidney/brain/vector: pBAC63.6; Site\_1: EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI methylase. Size selected DNA was cloned into the pBAC63.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies). "

BASE COUNT	226 a	130 c	94 g	140 t
ORIGIN				

Query Match	9.8%;	Score 26;	DB 12;	Length 590;
Best Local Similarity	100.0%;	Pred. No. 9.2;		

Matches	26;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	32	AAAAAATTAAAAACAGAAAAACAAA	57						
Db	375	AAAAAATTAAAAACAGAAAAACAAA	400						

RESULT	3
AA748563/C	
LOCUS	AA748563
DEFINITION	oa57c09.s1 NCL_GCAP.GCB1 Homo sapiens cDNA clone IMAGE:1309072 3'
	343 bp mRNA linear EST 27-JAN-1998

LOCUS	343 bp	mRNA	EST 27-JAN-1996
DEFINITION	AA574859.s1	NCI CGAP GCBI	Homo sapiens cDNA clone IMAGE:130972.3
	similar to SMYCH1_MOUSE	P40201	CHROMODOMAIN-HELICASE-DNA-BINDING PROTEIN CSD-1. [1]; contains element L1 repetitive element; ; mRNA sequence.

ACCESSION	AA/48563	GI:27885211
VERSION	AA748563.1	
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM	REFERENCE
Homo sapiens	1 (bases 1 to 343)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	
Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.	

**AUTHORS** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
**JOURNAL** Tumor Gene Index  
**COMMENT** Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Louis M. Staedt, M.D., Ph.D., David Allman,  
Ph.D., Gerald Marti, M.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome  
Clone distribution: NCI-CCAP clone distribution  
found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.linnl.gov/bdrrp/image/image.html](http://www.bio.linnl.gov/bdrrp/image/image.html)  
Insert Length: 1312 Std Error: 0.00  
Seq primer: -40m13 fwd. Et from Amersham  
High quality sequence stop: 282.

FEATURES  
source

```

"organism":"Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1309072"
/clone_1id="NCI CGAP GCBI"
/tissue_type="germinal center B cell"
/lab_host="RDH10R"
/note="vector: pV73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI. 1st strand
was prepared from human tonsillar cells enriched for

```

germinal center B cells by flow sorting (Dr. David Allman, provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CEBR). cDNA synthesis was primed with a Not I - oligo(dt) primer (5'-TGTTCACCAATCTCAAGTCGAGGCGCCCTCATTTTCTTTTCTTTT-3'). Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonalido."

BASE COUNT	62 a	69 c	35 g	177 t
ORIGIN				

Query Match	9.1%	Score 24	DB 9	Length 343
Best Local Similarity	100.0%	Pred. No. 54		
Matches 24	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY 157 AAAAGAAAAAGAGCATAGAAGA 180  
|||  
Db 52 AAAAGAAAAAGAGCATAGAAGA 29

RESULT 4  
LOCUS A2213624 375 bp DNA linear GSS 09-JUN-2000  
DEFINITION Sheared DNA-113F9.TF Sheared DNA Trypanosoma brucei genomic clone  
ACCESSION A2213624  
VERSION A2213624.1 GI:8431424  
KEYWORDS GSS.  
SOURCE Trypanosoma brucei.  
ORGANISM Trypanosoma brucei.  
REFERENCE Eukaryota: Euglenozoa; Kinetoplastida; Trypanosomatidae;  
Trypanosoma  
1 (bases 1 to 375)  
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gull,S., Suh,E., Malek,J., Fujii,C.,  
Gerard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,  
Fraser,C. and Adams,M.  
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat  
10.1 sheared DNA library  
JOURNAL Unpublished (1999)  
COMMENT Other-GSSs: Sheared DNA-113F9.TF  
Contact: Najib M. El-Sayed  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: nelsayed@tifg.org  
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared  
DNA library constructed at TIGR. Clones will be available for  
distribution through Research Genetics, Alabama, USA. Sheared DNA  
end sequences search page: <http://www.tifg.org/cdb/mbd/cdbdb/>.  
Seq primer: M13-Forward  
Class: Shotgun  
FEATURES  
Source  
Location/Qualifiers  
1..375  
/organism="Trypanosoma brucei"  
/strain="TREU927/4 GUTat 10.1"  
/db\_xref="taxon:5691"  
/clone\_lib="Sheared DNA-113F9"  
/clone\_1lb="Sheared DNA"  
/note="Vector: pUC18; Site\_1: SmaI; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically  
sheared to give a tight size distribution (approx 2 kb).  
The v + i method used for the library construction is  
described in detail in Smith, H.O. and Venter, J.C.  
(Making small insert libraries for whole genome shotgun  
sequencing projects. In Genome Sequencing: A Practical  
Approach, eds. M. Vaudin and B. Barrell, Oxford University  
Press, 1999)."  
BASE COUNT 94 a 68 c 70 g 143 t  
ORIGIN

Query Match 9.1%; Score 24; DB 12; Length 375;  
Best Local Similarity 100.0%; Pred No. 53;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 TAAGAAAGATTAAAGAAAAGA 195  
|||||  
Db 308 TAAGAAAGATTAAAGAAAAGA 285

RESULT 5  
LOCUS A1242163 430 bp mRNA linear EST 01-DEC-1998  
DEFINITION qh81f02.x1 Soares\_fetal\_liver\_spleen\_infls\_S1 Homo sapiens cDNA  
clone IMAGE:1851099 3' similar to contains element LTR5 repetitive  
element; mRNA sequence.  
ACCESSION A1242163  
VERSION A1242163.1 GI:3837560

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 430)  
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaps@mail.nih.gov  
This clone is available royalty-free through LNC; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 1274 Std Error: 0.00  
Seq primer: -400P from Glbco  
High quality sequence stop: 429.  
FEATURES  
Source  
Location/Qualifiers  
1..430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:1851099"  
/clone\_1lb="Soares\_fetal\_liver\_spleen\_infls\_S1"  
/sex="male"  
/dev\_stage="20 week post conception fetus"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)  
with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI;  
This is a subtracted version of the original Soares fetal  
liver spleen infls library. 1st strand cDNA was primed  
with a Pac I - oligo(dT) primer [5'  
AACTGAAAGATTAAATAGATCTTTTCTTTTCTTTT 3']  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Pac I and cloned into the Pac I  
and Eco RI sites of the modified pT73 vector. Library  
went through one round of normalization. Library  
constructed by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 111 a 89 c 50 g 180 t  
ORIGIN

Query Match 9.1%; Score 24; DB 9; Length 430;  
Best Local Similarity 100.0%; Pred No. 51;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 AAAAGAAAGAGGATAGAAAGA 180  
|||||  
Db 51 AAAAGAAAGAGGATAGAAAGA 28

RESULT 6  
LOCUS AA699918 456 bp mRNA linear EST 19-DEC-1997  
DEFINITION 2161f12.s1 Soares\_fetal\_liver\_spleen\_infls\_S1 Homo sapiens cDNA  
clone IMAGE:435311 3', mRNA sequence.  
ACCESSION AA699918  
VERSION AA699918.1 GI:2702861  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 456)  
AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,  
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin  
J., Moore,B., Schellenberg,K., Stepcoe,M., Tan,F., Theisling,B.,  
White,Y., Wylie,T., Waterston,R. and Wilson,R.  
TITLE WashU-NCI human EST Project  
JOURNAL Unpublished (1997)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (infoimage.lnl.gov) for further information.  
Seq primer: -40m13 fwd. ER from Amersham  
High quality sequence stop: 420.

Location/Qualifiers

## FEATURES

source

```

1. .485
/organism="Homo sapiens"
/db_xref="GDB:1335080"
/db_xref="taxon:9606"
/clone="IMAGE:435311"
/clone_id="Soares_fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
ACCTGGAAGATTAAATTAAGATCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

```

BASE COUNT  
ORIGIN

119 a 95 c 57 g 185 t

Query Match 9.1%; Score 24; DB 9; Length 456;  
Best Local Similarity 100.0%; Pred. No. 50;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 157 AAAAGAAAAAGAGTAAGAAAGA 180

Db 47 AAAAGAAAAAGAGATAGAAAGA 24

## RESULT 7

LOCUS

DEFINITION

AP642393 485 bp DNA linear GSS 08-JUL-1999  
RPC193-DpniI-28K15.TV RPC193-DpniI Trypanosoma brucei genomic clone

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Trypanosoma brucei.  
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Other\_GSSs: RPC193-DpniI-28K15.TV

Contact: Na'ib M. El-Sayed

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: neisayed@tigr.org

Clones and high density filters may be purchased from BACPAC

Resources (http://bacpac.med.bu.edu). BAC end sequences search

page: http://www.tigr.org/tdb/mdb/tbdb/.

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

FEATURES

source

```

1. .485
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="RPC193-DpniI-28K15"
/clone_id="RPC193-DpniI"
/notes="Vector: pBACE3.6; Site_1: Bam HI; Site_2: Bam HI;
constructed for The Institute for Genomic Research by
Bonni Zhao in Pieter de Jong's laboratory (Roswell Park
Cancer Institute, Buffalo, NY). Briefly, Trypanosoma
brucei TREU927/4 GUTat 10.1 agarose embedded DNA was
partially digested with a combination of Eco RI and Eco RI
methylase (RPC193-EcoRI segment) or Dpn II (RPC193-DpniI
segment). High molecular weight fragments were ligated in
pBACE3.6 vector digested with Eco RI or Bam HI,
respectively. The average insert size is 141 Kb. Total
coverage (both segments): > 90 x the haploid
non-mitochondrial genome."

```

BASE COUNT  
ORIGIN

172 a 91 c 68 g 153 t 1 others

Query Match 9.1%; Score 24; DB 12; Length 485;  
Best Local Similarity 100.0%; Pred. No. 50;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 172 TAAGAAAGATTAAAGAAAGA 195

Db 64 TAAGAAAGATTAAAGAAAGA 87

## RESULT 8

LOCUS

AO500417 539 bp DNA linear GSS 29-APR-1999  
V40B12 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Michael Snyder, Dept. of Mol. Cell. and Dev. Biology

Yale University

P.O. Box 208103, New Haven, CT 06520-8103, USA

Tel: 203 432 9949

Fax: 203 432 6161

Email: anuj.kumar@yale.edu

te of mTn-3xHA/lacZ insertion.

Seq primer: GGCTCTCTCTTGGGAAGTAC

Class: transposon-tagged.

Location/Qualifiers

1. .539

/organism="Saccharomyces cerevisiae"

/db\_xref="taxon:4932"

/clone\_id="mTn-3xHA/lacZ Insertion Library"

/lab\_host="E. coli"

/note="Vector: pHS56-Sal; A yeast genomic DNA library

(lacking mitochondrial DNA) was prepared in pHS56-Sal;

genomic DNA was size-fractionated (DNA of roughly 2-3 kb

in length) prior to cloning. This library was

subsequently mutagenized with a mTn-3xHA/lacZ

mitransposon containing lacZ, URA3, and tet resistance."

FEATURES

source

FEATURES



QY 29 ATATAAAATATATAACAGAAAAA 52  
 |||||||  
 Db 298 ATATAAAATATATAACAGAAAAA 275

RESULT 11  
 BM015597 677 bp mRNA linear EST 30-OCT-2001  
 LOCUS 603641828F1 NIH\_MGC\_87 Homo sapiens cDNA clone IMAGE:5417887 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BM015597  
 VERSION BM015597.1 GI:16529951  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 677)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: DCTD/DTP  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM12065 row: 9 column: 08  
 High quality sequence stop: 674.  
 Location/Qualifiers  
 1. 677  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5417887"  
 /clone\_1b="NIH\_MGC\_87"  
 /tissue\_type="mammary adenocarcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: breast; Vector: pCMV-SORT6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.383 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC library."

BASE COUNT 305 a 92 c 153 g 127 t

ORIGIN

Query Match 9.1%; Score 24; DB 10; length 677;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AAAAGAAAAAGAGATTAAGAAAGA 180  
 |||||||  
 Db 245 AAAAGAAAAAGAGATTAAGAAAGA 268

RESULT 12  
 BF239967 821 bp mRNA linear EST 14-NOV-2000  
 LOCUS 601905170F1 NIH\_MGC\_54 Homo sapiens cDNA clone IMAGE:4133129 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BF239967  
 VERSION BF239967.1 GI:11153890  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 821)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LUCM1033 row: k column: 18  
 High quality sequence stop: 562.  
 Location/Qualifiers  
 1. 821  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4133129"  
 /clone\_1b="NIH\_MGC\_54"  
 /tissue\_type="from chronic myelogenous leukemia"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);  
 Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggccataggcc  
 ); Double-stranded cDNA was prepared from cell line RNA.  
 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGGCCATATAGGCC-3' and 3' adaptor  
 sequence: 5'-ATCTAGAGCGCGCGCGCGCACAG-3' (BN-3',  
 where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and was constructed by Clontech  
 Laboratories (Palo Alto, CA)."

BASE COUNT 345 a 149 c 190 g 137 t

ORIGIN

Query Match 9.1%; Score 24; DB 10; length 821;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AAAAGAAAAAGAGATTAAGAAAGA 180  
 |||||||  
 Db 391 AAAAGAAAAAGAGATTAAGAAAGA 414

RESULT 13  
 AU039357 372 bp mRNA linear EST 29-MAR-1999  
 LOCUS AU039357 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
 DEFINITION AU039357 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium  
 ACCESSION AU039357  
 VERSION AU039357.1 GI:4008597  
 KEYWORDS EST.  
 SOURCE Dictyostelium discoideum.  
 ORGANISM Dictyostelium discoideum  
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 1 (bases 1 to 372)  
 Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,  
 Yoshino,R., Mittra,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H.,  
 Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.  
 The Dictyostelium development cDNA project: generation and  
 analysis of expressed sequence tags from the first-finger stage of  
 development  
 DNA Res. 5 (6), 335-340 (1998)  
 99156227  
 JOURNAL Contact: Hideko Urushihara  
 MEDLINE Institute of Biological Sciences  
 COMMENT University of Tsukuba  
 3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan  
 Email: d402h@tsukuba.ac.jp  
 PROJECT = "Dictyostelium discoideum cDNA project in Japan".  
 Location/Qualifiers  
 1. 372  
 /organism="Dictyostelium discoideum"



/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="SLH629"  
/clone\_11b="Dictyostellium discoideum SL (H. Drushihara)"  
/dev\_stage="sing"  
BASE COUNT 234 a 40 c 51 g 46 t 1 others  
ORIGIN

Query Match 8.7%; Score 23; DB 9; Length 372;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 AAAGAAAAGATATTAAGAAA 207  
|||||  
DB 252 AAAGAAAAGATATTAAGAAA 274

RESULT 14  
AA658426/c 393 bp mRNA linear EST 05-NOV-1997  
LOCUS  
DEFINITION nubl1.81 NCI\_CGAP\_P12 Homo sapiens cDNA clone IMAGE:1208396, mRNA  
sequence.  
ACCESSION AA658426 GI:2594580  
VERSION AA658426.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 393)  
NCI\_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: rgs@nci.nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquai, M.D.  
Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILMIL at:  
www.bio.illn.gov/db/ncp/image/image.html  
Seq primer: -40m13 fwd. fr from Amerham.

FEATURES  
SOURCE  
Location/Qualifiers  
1..393

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1208396"  
/clone\_11b="NCI\_CGAP\_P12"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: PAMPI0; Site 1: NotI; Site 2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated total cellular RNA obtained from 5,000-10,  
000 microdissected preneoplastic cells  
histologically-determined to be prostatic intraepithelial  
neoplasia 2 (PIN2) cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into PAMPI0 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Kitzman."

BASE COUNT 121 a 58 c 69 g 145 t  
ORIGIN

Query Match 8.7%; Score 23; DB 9; Length 393;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 41 AAACGAGAAAAAGAAATGAGA 63  
|||||  
DB 391 AAACGAGAAAAAGAAATGAGA 369

RESULT 15  
AL653205 683 bp mRNA linear EST 13-DEC-2001  
LOCUS  
DEFINITION AL653205 XGC-gastrula silurana tropicalis cDNA clone Tgas049n16.5',  
mRNA sequence.  
ACCESSION AL653205 GI:17664069  
VERSION  
KEYWORDS EST.  
SOURCE western clawed frog.  
ORGANISM silurana tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Silurana.  
1 (bases 1 to 683)  
Huckle, E., Taylor, R., Ashurst, J. L., Zorn, A. M. and Rogers, J.  
Sanger Xenopus tropicalis EST project 2001 (10\_2001)  
Unpublished (2001)  
Contact: Huckle E  
Sanger Centre  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: Tgas049n16.sp6  
Sequencing primer: SP6  
This sequence is from a Xenopus Gene Collection (XGC) library  
constructed by Aaron M. Zorn.

FEATURES  
SOURCE  
Location/Qualifiers  
1..683

/organism="Silurana tropicalis"  
/db\_xref="taxon:8364"  
/clone="Tgas049n16"  
/clone\_11b="XGC-gastrula"  
/dev\_stage="gastrula (stages 10.5-13 mixed)"  
/lab\_host="Escherichia coli XL1-blue"  
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA  
was oligo dT primed from sug of poly A+ RNA from stages  
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated  
into pCS107 with EcoRI at the 5' end and NotI at the 3'  
end."

BASE COUNT 225 a 113 c 110 g 234 t 1 others  
ORIGIN

Query Match 8.7%; Score 23; DB 9; Length 683;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 AATATTAACACAAAAAGAAAA 57  
|||||  
DB 517 AATATTAACACAAAAAGAAAA 539

Search completed: August 2, 2002, 22:42:01  
Job time: 30176 sec

---





TITLE  
JOURNAL  
COMMENT

Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupp, R., Seaman, S.,  
Severy, P., Spencer, B., Stange-Thumann, N., Stojanovic, N.,  
Strauss, N., Sudrmanian, A., Talam, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, R.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (27-JAN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www.seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L20703

Center clone name: 438\_D3

\* NOTE: This record contains 77 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1	728:	contig of 728 bp	in length
729	828:	gap of 100 bp	
829	1553:	contig of 725 bp	in length
1554	1653:	gap of 100 bp	
1654	2373:	contig of 720 bp	in length
2374	2473:	gap of 100 bp	
2474	3176:	contig of 703 bp	in length
3177	3276:	gap of 100 bp	
3277	3998:	contig of 722 bp	in length
3999	4098:	gap of 100 bp	
4099	4821:	contig of 723 bp	in length
4822	4921:	gap of 100 bp	
4923	5607:	contig of 686 bp	in length
5608	5707:	gap of 100 bp	
5708	6414:	contig of 707 bp	in length
6415	6514:	gap of 100 bp	
6515	7238:	contig of 724 bp	in length
7239	7338:	gap of 100 bp	
7339	8067:	contig of 729 bp	in length
8068	8167:	gap of 100 bp	
8168	8898:	contig of 731 bp	in length
8899	8998:	gap of 100 bp	
8999	9728:	contig of 730 bp	in length
9729	9828:	gap of 100 bp	
9829	10554:	contig of 726 bp	in length
10555	10654:	gap of 100 bp	
10655	11300:	contig of 646 bp	in length
11301	11400:	gap of 100 bp	
11401	12105:	contig of 705 bp	in length
12106	12205:	gap of 100 bp	
12206	12867:	contig of 662 bp	in length
12868	12967:	gap of 100 bp	
12968	13666:	contig of 699 bp	in length
13667	13766:	gap of 100 bp	
13767	14494:	contig of 728 bp	in length
14495	14594:	gap of 100 bp	
14595	15328:	contig of 734 bp	in length
15329	15428:	gap of 100 bp	

15429	16122:	contig of 694 bp	in length
16123	16222:	gap of 100 bp	
16223	16944:	contig of 722 bp	in length
16945	17044:	gap of 100 bp	
17045	17738:	contig of 694 bp	in length
17739	17838:	gap of 100 bp	
17839	18534:	contig of 696 bp	in length
18535	18634:	gap of 100 bp	
18635	19333:	contig of 699 bp	in length
19334	19433:	gap of 100 bp	
19434	20145:	contig of 712 bp	in length
20146	20245:	gap of 100 bp	
20246	20972:	contig of 727 bp	in length
20973	21072:	gap of 100 bp	
21073	21805:	contig of 733 bp	in length
21806	21905:	gap of 100 bp	
21906	22639:	contig of 734 bp	in length
22640	22739:	gap of 100 bp	
22740	23460:	contig of 721 bp	in length
23461	23560:	gap of 100 bp	
23561	24284:	contig of 724 bp	in length
24285	24384:	gap of 100 bp	
24385	25093:	contig of 715 bp	in length
25100	25199:	gap of 100 bp	
25200	325902:	contig of 703 bp	in length
25203	26002:	contig of 100 bp	
26003	26698:	contig of 696 bp	in length
26699	26798:	gap of 100 bp	
26799	27503:	contig of 705 bp	in length
27504	27603:	gap of 100 bp	
27604	28337:	contig of 734 bp	in length
28338	28437:	gap of 100 bp	
28438	29161:	contig of 724 bp	in length
29162	29261:	gap of 100 bp	
29262	29984:	contig of 723 bp	in length
29985	30084:	gap of 100 bp	
30085	30784:	contig of 700 bp	in length
30785	30884:	gap of 100 bp	
30885	31509:	contig of 625 bp	in length
31510	31609:	gap of 100 bp	
31610	32316:	contig of 707 bp	in length
32317	32416:	gap of 100 bp	
32417	33135:	contig of 719 bp	in length
33136	33235:	gap of 100 bp	
33236	33944:	contig of 709 bp	in length
33945	34044:	gap of 100 bp	
34045	34766:	contig of 722 bp	in length
34767	34866:	gap of 100 bp	
34867	35351:	contig of 685 bp	in length
35352	35651:	gap of 100 bp	
35653	36388:	contig of 737 bp	in length
36389	36488:	gap of 100 bp	
36489	37208:	contig of 720 bp	in length
37209	37308:	gap of 100 bp	
37309	38035:	contig of 727 bp	in length
38036	38135:	gap of 100 bp	
38136	38858:	contig of 723 bp	in length
38859	38958:	gap of 100 bp	
38959	39684:	contig of 726 bp	in length
39685	39784:	gap of 100 bp	
39785	40482:	contig of 698 bp	in length
40483	40582:	gap of 100 bp	
40583	41286:	contig of 704 bp	in length
41287	41386:	gap of 100 bp	
41387	42090:	contig of 704 bp	in length
42091	42190:	gap of 100 bp	
42191	42911:	contig of 721 bp	in length
42912	43011:	gap of 100 bp	
43012	43740:	contig of 729 bp	in length
43741	43840:	gap of 100 bp	
43841	44567:	contig of 727 bp	in length
44568	44667:	gap of 100 bp	
44668	45371:	contig of 704 bp	in length

```
* 45372 45471: gap of 100 bp
* 45472 46199: contig of 728 bp in length
* 46200 46299: gap of 100 bp
* 46300 47029: contig of 730 bp in length
* 47030 47129: gap of 100 bp
* 47130 47847: contig of 718 bp in length
* 47848 47947: gap of 100 bp
* 47948 48664: contig of 717 bp in length
* 48665 48764: gap of 100 bp
* 48765 49486: contig of 722 bp in length
* 49487 49586: gap of 100 bp
* 49587 50310: contig of 724 bp in length
* 50311 50410: gap of 100 bp
* 50411 51113: contig of 703 bp in length
* 51114 51213: gap of 100 bp
* 51214 51909: contig of 696 bp in length
* 51910 52009: gap of 100 bp
* 52010 52708: contig of 699 bp in length
* 52709 52808: gap of 100 bp
* 52809 53540: contig of 732 bp in length
* 53541 53640: gap of 100 bp
* 53641 54377: contig of 737 bp in length
* 54378 54477: gap of 100 bp
* 54478 55200: contig of 723 bp in length
* 55201 55300: gap of 100 bp
* 55301 56030: contig of 730 bp in length
* 56031 56130: gap of 100 bp
```

## alignment\_scores:

```
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

## alignment\_block:

```
US-08-973-363-14 x AC108422/rev ..
```

```
Align seg 1/1 to reverse of: AC108422 from: 1 to: 62575
```

```
34 GlycyluysArgGluThrLysGluLys 42
|||||
5562 GGGGAAAAGCAGACAGACAAAAGAAAA 5536
```

```
seq_name: gb_pl:ATF1C12
```

```
seq_documentation_block:
```

```
LOCUS ATF1C12 111945 bp DNA linear PLN 20-SEP-1999
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F1C12 (ESSA
project).
```

```
ACCESSION AL022224
VERSION AL022224.1 GI:3059018
```

```
KEYWORDS
SOURCE
ORGANISM
```

```
Arabidopsis thaliana
thale cress.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
```

## REFERENCE

## AUTHORS

Bevan, M., Terry, N., Ardiles, W., Buyssehaert, C., Dasseville, R., De  
Clerck, R., De Keyser, A., Neyt, P., Rouze, P., Van Den Daele, H.,  
Villarreal, R., Gielens, J., Van Montagu, M., Bancroft, I., Mewes, H. W.,  
Mayer, K. F. X., Lemcke, K. and Schueller, C.  
unpublished

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (20-SEP-1999) MIPS, at the Max-Planck-Institut fuer  
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:  
schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project  
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge  
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,  
E-mail: michael.bevan@bbsrc.ac.uk  
On Apr 18, 1998 this sequence version replaced gi:2982425.  
Information on performance of analysis and a more detailed

## FEATURES

## source

annotation of this entry and other sequences of chromosomes 3, 4  
and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

Location/Qualifiers

1. 111945 "Arabidopsis thaliana"

/organism="Arabidopsis thaliana"

/variety="Columbia"

/db\_xref="taxon:3702"

/chromosome="4"

14368..16954

/gene="F1C12.60"

/number=1

14368..18193

/gene="F1C12.60"

join(14368..16954,17006..17764,17841..18193)

/gene="F1C12.60"

/note="similarity to cf-2.2, Lycopersicon

pimpinellifolium, PIR:T10515

contains EST gb:Z38045, Z46532"

/codon\_start=1

/product="leucine rich repeat-like protein"

/protein\_id="CA18239.1"

/db\_xref="GI:2982431"

/translation="MPPVILLLFLICPSGLGPGIINNDQTLLEPKSVYTNPED

DELRQWNSDINICSWTGVYCDNTGLFRVYALNLVTLGLGTSISPFGRFDNLIDIL

SSNNLVGPIPTALSNLTSLESLFENQNLGELIPSQLGLVNRSLRIDNDELVDIP

ETLGNLVNLOMLALASCRLTGPIPSOLGRIVRQSLIDQNYLEGIIPALGNCSDLT

VPTAAENMLNGITIPALGRLNLEIINLANNSTLGEIPSQLGMSQLOVLSKANOLO

GLTPKSLADLGNLQTLIDISANNLTGELIPPEPMMNSQLDLVLANNNLSGSLTCS

NNTLEOLVLSGTLGSEIPEVLSKQSLKQDLSNNSLAGSLPEALFELVELDTLISN

NNTLEGLTSPSISNLTNLQMLVLYHNHLEKLEKETLSLRKLELVLEKRSGEIIPQ

EIGNCTSLKIDMGHFGELIPSPISGRKEKLLHLRQNEVLGGIPASLGLGSLP

LDLADNQLSGSISPSFGLKGLDLMLVNSLQGLNPLDSLSLRNLTIRINLSHRING

TIHPLCGSSSYLSFDVYNNNGFEDEIPELGNSQNLDRLRKKQNLGKIPWTLKRS

LSLLDMSNALNGTITPOLVLYLCKLTHIDLNNFLSGPIPPMTGLKSLQGLKTSNO

FVBSLPTELPNCKLLVLSIDGNSLGSIPQELNGALVNLIDKNSGSLPQAMG

KSKLEYLRSLRNSLTGEIPEVETGQDLQSLADLSYNNFTGDIPESTIGLSKLETL

LSHNQTLGEPVSGVGMKSLGYLVNFSNNLGGKLRQFSPMPADSLFGLNGLGSPLS

RCNRVRTISALFTAIIGLIIIVIALFEFRDHFKEKVGHSAYTSSSSSQARHKPLER

NGASKDIRMEDMEATHNLSEEFNIGSGSGGVYAELENGETVAVKKILMDKDLMS

NKFSREVRITGRIRHRLVKMGSSSEGLNLTLYEYMKKSLDNNLEHEDKPELE

KKKKLLDMERAKRLIAGLQAGVYLYHDCVPPYVHDIIDSSNLTSSNLEAHKCDPGL

AKVLTENCIDNTDNTWFCSTGITYAPETAYSLKATKESDVSIMGIVLMEIYTKMPT

DSVFGAKMDVMVRVETHELPAVGASARDKLIDPKLKLPLPEEDAACVLEIALQCTKTS

POERPSSRQACDLSLHVYNNRTAGYRKL"

16955..17005

/gene="F1C12.60"

/number=1

17006..17764

/gene="F1C12.60"

/number=2

17765..17840

/gene="F1C12.60"

/number=2

17841..18193

/gene="F1C12.60"

/number=3

complement(join(19211..19437,19558..19648))

/gene="F1C12.70"

complement(19211..19437)

/gene="F1C12.70"

/number=1

19211..19648

/gene="F1C12.70"

complement(join(19211..19437,19558..19648))

/note="predicted

contains EST gb:T21779, R90402, Aa712327, R64867,

AA042326, Aa712523, Aa712771, Aa605387, F13839"

/codon\_start=1

/product="putative protein"

/protein\_id="CA18240.1"

/db\_xref="GI:2982432"

/db\_xref="SPTRMBL:065429"

```

/translating="NPISATVYCALIGLGTOMYSNLRKLPYMRPHEHYGMLGAY
FANOLVKMDVLEKEDDVLAKARANRRRYFQIYTPILVYGLSAGSILLFPRKIGDV
FCV"
complement(19438..19557)
/gene="F1C12.70"
/number=1
complement(19558..19648)
/gene="F1C12.70"
/number=2
21131..22045
/gene="F1C12.80"
/number=1
21131..25119
/gene="F1C12.80"
join(21131..22045,22134..22261,22328..24208,24317..24665,
24743..24890,24974..25119)
/gene="F1C12.80"
/notes="similarity to mature-parasite-infected erythrocyte
surface antigen ME5A, Plasmodium falciparum, PIR2:A45605"
/codon_start=1
/product="glu-rich protein"
/protein_id="CA18241.1"
/db_xref="GI:2982433"
/translation="MASSPEVMSQOTFNCVLSRPHERTQORONTNNKKNNNNN
VNINIVAFKKNLNFVRDPLFNCSSVDDVDNDSTKAPPCSSKSSVDNRSLKPE
ESROVPSSSSLGSDSPNSGPPNSPMAASITVQIAPPTQCPSPSSNLSIDSR
TSSKSNLRNLSSELSSEYKESQMLDPIERCNNEEERETECAPPRPESGKREGRV
MOTIKRINSEETITNDGSSGNDNKEVQTTKAPFVACSPRIHQMLADLV
QMRREKDLACLREKCYKFTNKGKOSTLRTRECLATGRRRSKTSKSDS
NRSYSGVSHMLREKTKANSENTEGASTSHSTRGRIMKDRHAKTEKVALTE
KSGVLEKEVSAENAKGVLSERNGKLKLEGTQREVEKREKREKREKREK
ISSELEKGEKREKREKREKREKREKREKREKREKREKREKREKREKREK
ESGNTVERKENDVLEATRRILVESAESSTTSKETMTRCVAKVKKKKRE
VLESNKREKEDGNTKPGQVTDQDCLRPAIEKVALQETVKSOLKPEKRSRE
ERRELEESTISGIAEKALKEKREKREKREKREKREKREKREKREKREKREK
RISNVEVERSIDTSMITVYVAVGDRVIMQKREKREKREKREKREKREKREK
KSDVIEDKNNQPAVIGHSGKREKREKREKREKREKREKREKREKREKREK
EKSVSHGSRKREKREKREKREKREKREKREKREKREKREKREKREKREK
ESAKENSSQODESTHNPMDKGIKEPEDESKKREKREKREKREKREKREK
DONEMEEDEEEDYFNGDDWDHDSIRERSYMEKREKREKREKREKREK
IRRTVSNFLTSLRQKIDNLTITRVOSSHVLANIEGDEVEEWEVCSAFLPE
NTEEEEPKTELEAPSDVCSOASASSSTMSNFRDODIDNDENPTSLPELPT
NOSTODMOTLIDLEKEMOLOREMLELRNTVKSCIMDLOKSVTODLSRSQSAEQ
RDPPKDPDKKCKCEMPVDLSLYRFCLYIRNKSPTLK"
22046..22133
/gene="F1C12.80"
/number=1
22134..22261
/gene="F1C12.80"
/number=2
22262..22327
/gene="F1C12.80"
/number=2
22328..24208
/gene="F1C12.80"
/number=3
24209..24316
/gene="F1C12.80"
/number=3
24317..24665
/gene="F1C12.80"
/number=4
24666..24742
/gene="F1C12.80"
/number=4
24743..24890
/gene="F1C12.80"
/number=5
24891..24973
/gene="F1C12.80"
/number=5
24974..25119

```

```

/gene="F1C12.80"
/number=6
26862..26897
/notes="36bp TTC tandem repeat"
26958..28126
/gene="F1C12.90"
/number=1
26958..28942
/gene="F1C12.90"
join(26958..28126,28597..28942)
/gene="F1C12.90"
/notes="similarity to gene f4p9.34 chromosome II BAC f4p9,
Arabidopsis thaliana
contains EST gb:237280, 237281"
/codon_start=1
/product="putative protein"
/protein_id="CA18242.1"
/db_xref="GI:2982434"
/translation="MAVVKREKONTKDKLLVGYINFSAEKLTPMALLVLTATL
LPPIPSFSLSTSDPFCISRRSSAVPLNTTYVESSSPSPKENDRVLDGVYKR
TFPGYSAAYNFVMSAYRGVNSFAVIGLSKRLHYGHPISPCRWVSLDPTQDIS
TTEGFLTDMGCGRLTYTYVYVCTFSSISAVNQNGGTLILHATGDDPLNLTDSIS
VTEPKSYVDLNSYRKRRKRYDYLVGSSGTXGNSIPORVYAVHVRPGRSHF
VLADAGIHEEYFEVILKWTGKRGVTLIDTDPREDCYHNOGPIYNDLHFRPMT
KMFPEFDDEFLHVPVETISSMESLEYSQFTIDKMSRRCYSGSDPATRYKKM

```

alignment\_scores:

Quality:	9.00	Length:	9
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment\_block:

US-08-973-363-14 x ATF1C12 ..

Align seg 1/1 to: ATF1C12 from: 1 to: 111945

34 G1G1uLysArgGluThrLysGluLys 42

|||||

22605 GCGCAAAAGCGAGACGACGAGAGAA 22631

seq\_name: gb\_hhg:AC094613

seq\_documentation\_block:

LOCUS AC094613 173449 bp DNA linear HTG 20-DEC-2001

DEFINITION Rattus norvegicus clone CH230-5b2, \*\*\* SEQUENCING IN PROGRESS \*\*\*

ACCESSION AC094613

VERSION AC094613.2 GI:17941384

KEYWORDS HTG; HNGS; PHASEL.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 173449)

AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarta,J., Benton,J., Blimege,K., Blankenburg,K., Bonnin,D., Bouck,T., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dachorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Diaper,H., Denn,A.L., Ding,Y., Dinh,H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garra,N., Gill,R., Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,R., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Hollway,C., Hollins,B., Honsi,F., Howard,S., Huber,J., Huljk,S., Hume,J.,

Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,  
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J.,  
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,  
Lewis, L., Li, J., Li, Z., Lichteberg, O., Lieu, C., Liu, J., Liu, W.,  
Loui, S., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,  
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,  
Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M.,  
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,  
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,  
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S.,  
Ogih, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,  
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,  
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,  
Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshbari, N.,  
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,  
Stang, H., Sutton, A., Svatek, A., Tabori, P., Tamerisa, A., Tamerisa, K.,  
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,  
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,  
Washington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,  
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

Direct Submission

Unpublished

2 (bases 1 to 173449)

Worley, K.C.

Direct Submission

Submitted (15-SEP-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:15624448.

## COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GBAC

Center clone name: CH230-582

----- Summary Statistics

Assembly program: Phrap; version 0.990329first call to  
findPhrapList

Consensus quality: 130278 bases at least Q40

Consensus quality: 139554 bases at least Q30

Consensus quality: 147493 bases at least Q20

Estimated insert size: 137875; sum-of-coverage estimation

Quality coverage: 0x in Q20 bases; agrose-fp estimation

Quality coverage: 1.9x in Q20 bases; sum-of-coverage estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 66 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 6988: contig of 6988 bp in length

6989 7088: gap of unknown length

7089 12288: contig of 5200 bp in length

12289 12388: gap of unknown length

12389 21337: contig of 8939 bp in length

21338 21437: gap of unknown length

21438 27590: contig of 6163 bp in length

27591 27690: gap of unknown length

27691 33366: contig of 5676 bp in length

33367 33466: gap of unknown length

33467 38762: contig of 5296 bp in length

38763 38862: gap of unknown length

38863 42843: contig of 3981 bp in length

42844 42943: gap of unknown length



```
* 138625 138724: gap of unknown length
* 138725 140704: contig of 1980 bp in length
* 140705 140804: gap of unknown length
* 140805 142710: contig of 1996 bp in length
* 142711 142810: gap of unknown length
* 142811 145837: contig of 3027 bp in length
* 145838 145937: gap of unknown length
* 145938 147143: contig of 1208 bp in length
* 147146 147245: gap of unknown length
* 147246 148641: contig of 1396 bp in length
* 148642 148741: gap of unknown length
* 148742 150007: contig of 1266 bp in length
* 150008 150107: gap of unknown length
* 150108 151754: contig of 1647 bp in length
* 151755 151854: gap of unknown length
* 151855 153646: contig of 1792 bp in length
* 153647 153746: gap of unknown length
* 153747 155149: contig of 1403 bp in length
* 155150 155249: gap of unknown length
* 155250 156827: contig of 1578 bp in length
* 156828 156927: gap of unknown length
* 156928 158567: contig of 1640 bp in length
* 158568 158667: gap of unknown length
* 158668 160082: contig of 1415 bp in length
* 160083 160182: gap of unknown length

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-08-973-363-14 x AC094613/rev ..
  Align seg 1/1 to reverse of: AC094613 from: 1 to: 173449
  34 GlycylutysarglutThlysglutys 42
  ||||||||||||||||||||||||||||
  145411 GCGGAAAAAGGACCAAGAGAAA 145385

seq_name: gb_pl:ATCHRIV52

seq_documentation_block:
  LOCUS ATCHRIV52 198427 bp DNA linear PLN 16-MAR-2000
  DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 52.
  ACCESSION AL161552
  VERSION AL161552.2 GI:7268789
  KEYWORDS
  SOURCE
  ORGANISM Arabidopsis thaliana
  chae cress.
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
  1 (bases 67424 to 179368)
  Terry,N., Ardiles,W., Buysshaert,C., Dasseville,R., De Clerck,R.,
  De Keyser,A., Neyt,P., Rouze,P., Van Den Daele,H., Villarroel,R.,
  Gielens,J., Van Montagu,M., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  Unpublished
  2 (bases 69232 to 69818)
  Voelckert,G., Grymoprez,B., Voelt,M., Robben,J., Mewes,H.W.,
  Lemcke,K. and Mayer,K.F.X.
  Unpublished
  3 (bases 178975 to 187911)
  Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  Unpublished
  4 (bases 187649 to 198427)
  Pohl,T., Weizenegeer,T., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
  Unpublished
  5 (bases 1 to 82704)
  Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
  Mayer,K.F.X.
  Unpublished
  6 (bases 1 to 198427)
```

```

AUTHORS EU Arabidopsis sequencing project.
TITLE Direct Submission
JOURNAL Submitted (10-MAR-2000) MIPS at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de
ProjectCoordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV51 at the 5' end and an
overlap with ATCHRIV53 at the 3' end.
FEATURES
source
  1..198427
  /organism="Arabidopsis thaliana"
  /variety="Columbia"
  /db_xref="taxon:3702"
  /chromosome="4"
  7547..8329
  /gene="At4g19950"
  /number=1
  join(7547..8329,8375..8512)
  /gene="At4g19950"
  /note="similarity to p48 eggshell protein gene,
  Schistosoma mansoni, P48CHX:645484
  contains EST gb:AI994604.1, R30520, AI992766.1, T20423,
  AV112864, H76323"
  /codon_start=1
  /product="putative protein"
  /protein_id="CAB78995.1"
  /db_xref="GI:7268790"
  /translation="MDLAPEDLFNKRGLIRESTSIPOYSKTEPYLTLLTLPFSLF
  AILASHLFPOLAIQIDTPQADQSLQHWTVLVFQCYIIFLFAFSLSTAAVF
  TVASLYGKPSSESTSAIPVLRKLFETFLWASLMLATVGLVLEIPLVLAIDL
  QNVLAIFSLVVIYFLVHVHVTALMHLASVSVLEIPLGLAMKSYELDKGTL
  MGSWVFITLVGCFIAGVAVVAGDGDGIFARIVAGGFLSVFYVCKSEFHQEI
  DKSLDHLGIGYLGREYVPLKSNQMNENEV"
  7547..8512
  /gene="At4g19950"
  8330..8374
  /gene="At4g19950"
  /number=1
  8375..8512
  /gene="At4g19950"
  /number=2
  11377..11526
  /gene="At4g19960"
  /number=1
  11377..14567
  /gene="At4g19960"
  join(11377..11526,11621..11846,11925..12170,12258..12332,
  12383..12643,12687..12887,12986..13288,13407..13929,
  14024..14567)
  /gene="At4g19960"
  /note="similarity to putative potassium transporter AtKTP2
  & AtKTP1. Arabidopsis thaliana, Patchy:G238469 &
  Patchy:G2384671
  contains EST gb:AI997834.1"
  /codon_start=1
  /product="potassium transporter-like protein"
  /protein_id="CAB78996.1"
  /db_xref="GI:7268791"
  /translation="MAERVFASVPEGENTFERENGAMMELEQKLPQDPDEANKLN
  KTRKGLSMMLLRLRSLQSLVIGDGLSPVIVYINFPQIDSDLYALSLIL
  ISLLPLIKIVCYCKRNGGGGLATVLSLCRIAKRALLPQNRSDDELTVTSRT
  VSAQSPAKTKRKLEKEMRKRALVYVLTGCMRGILTPAISVIGIVNNPK
  MSGNHLFTLSTDSIVLVAIVLILGLSMQHGDKVGMFLPAIVLIMPLTIGATG
  WYNCKDFTSVKASPLVYIVLIFKRRHDSWISLGGILSTAITASKCYASIPUNAGTEA
  LKADIIVPELLAIQATFFVFCLLAVAGCOAAVLIHKHRYQDAFGFASIPSVNIP
  MFIATGAIVGSOATISGTVSKQAVAHGCFPRVAVITKSKRGLQICPDIMIL
  MGCIAVTASFKKOSQIGNAVGMITTSYKKNYFSQMTAVVLAIVTLTLMIMIL
  VMGHVILVILFTPLSFVELSYFSAVIFKIDEGWVPLITAAISLWMSVWYATVVK
```





## alignment\_block:

US-08-973-363-14 x AF008215 ..

Align seg 1/1 to: AF008215 from: 1 to: 2176

17 LysgluansgluLysaspGlu 24

1734 AAGAGAAATGAAGAGAAAGACGAA 1757

seq\_name: gb\_pl:AB038492

seq\_documentation\_block:

LOCUS AB038492 2637 bp mRNA linear PLN 15-AUG-2001  
DEFINITION Atriplex gmelini AgNHX1 mRNA for Na/H antiporter Nhx1, complete

cds:

ACCESSION AB038492 GI:9857313

VERSION Na/H antiporter Nhx1.

KEYWORDS Atriplex gmelini cDNA to mRNA.

SOURCE ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Caryophyllales; Caryophyllales; Chenopodiaceae; Atriplex.

REFERENCE 1 (bases 1 to 2637)

Hamada, A., Shono, M., Xia, T., Ohta, M., Hayashi, Y., Tanaka, A. and

AUTHORS Hayakawa, T.

Isolation and characterization of a Na/H antiporter gene from the

halophyte Atriplex gmelini

plant molecular biology. 46 (1), 35-42 (2001)

2 (bases 1 to 2637)

Hayakawa, T., Hamada, A. and Tanaka, A.

Direct Submission

Submitted (17-FEB-2000) Takahiko Hayakawa, Plantech Research

Institute, Research Center: 1000 Kamoshida-cho, Aoba-ku, Yokohama,

Kanagawa 227-0033, Japan (E-mail: pt10012@cc.m-kagaku.co.jp,

Tel: +81-45-963-3520, Fax: +81-45-962-7492)

FEATURES

Location/Qualifiers

1..2637

/organism="Atriplex gmelini"

/db\_xref="taxon:118074"

682..2349

/gene="AgNHX1"

682..2349

/gene="AgNHX1"

/codon\_start=1

/product="Na/H antiporter Nhx1"

/protein\_id="BAH11940.1"

/db\_xref="GI:9857314"

/translation="MMSQSLSSLSGKMDLTTSDHNVVSMNLEVALLCGIVIGHL

ENRMNNEITALLIGLAVVILISGKSSHLVSEDLFFIYLLPPIFNAQFOV

KKKOFNRFITIVLFGAVGLVFTIISGALSIFKLDIGLELADYLAIGAFAT

DSVCTLOVNODEPILLYLVFGEVNDATSVLENATOSFDLIDIRIALQFNGN

FLYLFASITLIGAFGLISAYIKIKRGRHSMDREVALMMLAYISYLAELFYISG

LITVPCGIVASHYTHNTTSSRYTKKAFATLSVAVFLELYGMALDIETKRF

VSDSPGISVAVSSILGLVAVGRAAFVFLSWMLNKAQSSEKVFENQIYIWMAGL

MGAVSMALAYNQFRTSGHTOLEGNALIMTSTISVLFSTVFGILTKPLIMELPOP

KHFTSCSTVSDGSPKSYSLPLEGMDYEVNGNHDTEPRTITVSPSSIMRLN

APLTHVHYHWRKFDSEFMRPVFGRGFEVFGSPREGSTNNLVDR"

polysite /note="30 a nucleotides"

BASE COUNT 685 a 509 c 530 g 913 t

ORIGIN

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x AB038492/rev ..

Align seg 1/1 to reverse of: AB038492 from: 1 to: 2637

37 ArggluThrltySGluLysGluasn 44

322 AGAGAACGAAAGGAAAAAGAAAAC 299

seq\_name: gb\_v1:CASNS1

seq\_documentation\_block:

LOCUS CASNS1 2711 bp DNA linear VRL 02-AUG-1993  
DEFINITION Mouse Cas NS-1 retrovirus gag-*onc* fusion protein (v-cbl) gene, 3'

end.

ACCESSION J04169 GI:323269

VERSION gag-*onc* fusion protein: viral oncogene.

KEYWORDS CAS NS-1 retrovirus DNA, isolated from mouse (strain NRS/N)

previously infected with Cas-Br-M virus.

SOURCE Cas NS-1 murine leukemia virus

ORGANISM Viruses; Retroviridae; Gammaretrovirus.

REFERENCE 1 (bases 1 to 2711)

Langdon, W.Y., Hartley, J.W., Klincken, S.P., Russett, S.K. and

Morse, H.C. III.

Induces early B-lineage lymphomas

Proc. Natl. Acad. Sci. U.S.A. 86, 1168-1172 (1989)

Draft entry and computer-readable sequence for [1] kindly submitted

by W.Y. Langdon, 02-MAR-1989. Cas NS-1 is an acutely transforming

murine retrovirus that induces pre-B and pro-B cell lymphomas. It

was generated from the ecotropic Cas-Br-M virus by sequential

recombinations with endogenous retroviral sequences and a cellular

oncogene. The position of the oncogene junction was determined to

be in the gag gene for p10 at a position that removed the terminal

24 gag-encoded amino acids. The gag-*onc* fusion protein appears to

be responsible for fibroblast and pre-B cell transformation. The

authors call this oncogene v-cbl for Castlas B-lineage lymphoma.

FEATURES

Location/Qualifiers

1..2711

/organism="Cas NS-1 murine leukemia virus"

/db\_xref="taxon:11793"

&lt;1..1609

/note="gag-*onc* fusion protein"

/codon\_start=2

/protein\_id="AAA42885.1"

/db\_xref="GI:323270"

/translation="LERLKLSQVHPYDEPDEPQOETNVSMFISQAPDPIGRLEBLE

DKSKTGLVRAPEAKIFNKRREPERREIRKRETEKEPRRAEDOEORKEPRRRH

REMSKLATVYIGRODROGGERRRPOLDKOCATYKRTGHNAAAGGCRPSRSP

GSTPSLAERAPDPPYLAAGNVAKSSGAGGSGGAGGLIGLMDADQPPHHNHL

SPHPCTVDRKWEKWCMLDKVVRILCQNPVNAIKSPYIIDLDPYOHILKTVLSR

YEGMAETLGENEYFVFMENIMKRTQITSLFEKGRMEENESQPRRLITKLSIFS

HMAELKGTSGSLFGDPTFRITKADAEFWKAEKRTVPRKSPRQALHEVNPIS

GLEAMAKSTIOTCDYISVPEPFDIETPLFQWSSLINMNSIANTHGVAFLTYD

EKARLQKTIHRGSIYFRUSCTRLOMAGIYVTAAGNLTOTIPRKNPLFOALIDGFR

EGRTLPDRNPPDITGLCEPYPHS"

BASE COUNT 752 a 688 c 688 g 583 t

ORIGIN 1 bp upstream of XbaI site.

alignment\_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x CASNS1 ..

Align seg 1/1 to: CASNS1 from: 1 to: 2711

36 LysArggluThrltySGluLysGlu 43

227 AAGAGACAAACAGAAAGAAAAGAA 250

seq\_name: gb\_in:010414

seq\_documentation block:

LOCUS 010414 28990 bp DNA linear INV 05-OCT-2001  
DEFINITION Caenorhabditis elegans cosmid F42A10, complete sequence.  
ACCESSION 010414  
VERSION 010414.1 GI:500733  
KEYWORDS HTG.  
SOURCE  
ORGANISM Caenorhabditis elegans.  
Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida;  
Rhabditidae; Rhabditidae; Peloderae; Caenorhabditis.

REFERENCE  
AUTHORS 1 (bases 1 to 28990)  
TITLE The C. elegans Sequencing Consortium.  
JOURNAL Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2018 (1998)

REFERENCE  
AUTHORS 2 (bases 1 to 28990)  
TITLE Latreille, P.

REFERENCE  
AUTHORS 3 (bases 1 to 28990)  
TITLE The sequence of C. elegans cosmid F42A10

REFERENCE  
AUTHORS 4 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 5 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 6 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 7 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 8 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 9 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 10 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 11 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 12 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 13 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 14 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 15 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 16 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 17 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 18 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 19 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 20 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 21 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 22 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 23 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 24 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 25 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 26 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 27 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 28 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 29 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 30 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 31 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 32 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 33 (bases 1 to 28990)  
TITLE Direct Submission

REFERENCE  
AUTHORS 34 (bases 1 to 28990)  
TITLE Direct Submission

NOTICE: This sequence may not be the entire insert of this clone.  
It may be shorter because we only sequence overlapping sections  
once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted: all  
regions were double stranded, sequenced with an alternate chemistry  
or covered by high quality data (i.e., phred quality >= 30); an  
attempt was made to resolve all sequencing problems, such as  
compressions and repeats; all regions were covered by sequence from  
more than one ml3 subclone.

NEIGHBORING COSMID INFORMATION

The 5' cosmid is Y4269A, 500 bp overlap; 3' cosmid is C23610, 200 bp  
overlap. Actual start of this cosmid is at base position 7 of  
CEL42A10; actual end is at 7862 of CELC23610.

## NOTES:

Coding sequences below are predicted from computer analysis, using  
the program GeneFindor (P. Green and L. Hillier, ms in preparation).

## FEATURES

## source

1. 28990

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db\_xref="taxon:6239"

/chromosome="III"

/clone="F42A10"

/complement(699..3324)

/gene="F42A10.3"

/complement(101n(699..836,1364..1441,2028..2514,

3203..3324))

/gene="F42A10.3"

/note="exon 2 similar to small region of

phenylethanolamine N-methyltransferase, B0303.4 and

C1489.4; coded for by the following C. elegans cDNAs:

cm1298, yk444n2.5, B1174689, B1175667"

/codon\_start=1

/product="Hypothetical protein F42A10.3"

/protein\_id="AA19074.2"

/db\_xref="GI:15617835"

/translation="MSTLIDPTTSSDAQVTLPEYVKKFDPPEYLOFYSRDAI

EDGRVSLFALPFAQLMOTMRPSESLDLDIGAGPYSALCPADVAARVHLSPFV

DRNLDVLRKIRHETIDWPVTKVNRREGPPPTPOVCITVEEKARGVSGGSLF

ADVHMDVPELAKGVDIIVFTLESACNRYEYCKVTMMHHLSSGRLVLSV

LEDDEVNSGRKTFIHLNLSQIIDLAVSLGVANAKVYVLDGEGMTFMATKL"

4374..4454

/product="tRNA-Ser"

4561..4645

/product="tRNA-Ser"

5475..5939

/gene="F42A10.2"

/join(5475..5696,5827..6042,6389..6607,6684..6818,

6866..7000,7076..7378,7422..7571,7673..7760,8233..8503,

8615..8789,9889..9939)

/gene="F42A10.2"

/note="similar to ezrin (contains a domain found in band

4.1, ezrin, moesin, radixin and talin); contains

similarity to Pfam domain PF00373 (Band\_41), Score=234.3,

E-value=4.1e-75, N=1; PF00769 (ERM), Score=489.2,

E-value=1e-143, N=1; coded for by the following C. elegans

cDNAs: yk34h5.3, yk34h5.5, yk52e5.5, yk214b9.3, yk214b9.5,

yk406f4.5"

/codon\_start=1

/product="Hypothetical protein F42A10.2a"

/protein\_id="AA19073.1"

/db\_xref="GI:500735"

/translation="MGGSTFSPNPKPRIVAKVSTNADLEKIVKMGRLHFEAVC

RIGIEETVREGLPTKKNIPOMLONDKTGGDIOKDTSGSTNPLFLKRYEDV

EPRLIDARHFLPQIRPRLALSNMJCSPASVLAIVADMGDTESGVLTD

KHPKSVTDQIDMSADMRIRKMSKACGSKREKESLIRVADMDLMTGLTPT

FRNKETDLHGSLDAGLITKGVNRTTPPFSVSELTNIOKNNKFKMTYKVKSTS

FKREQDEKERYENKOKVEVLTLEKRAQENIRAEKANDQAEARASHEFTLMYK

KSEVAECENRSLNNKSEALIRERKAREAEITLAKQMSLADVSLDANKRMY

SSSALITDIPNMHOVSGHSGHPGSPGSGFQLOSLPTHHPQWMSLSLMPSSA

FSSPISNODPMSHOQPOOMSONTOAHQOOTOVONOPPYGSGCMIIIPPDSSISQI

FEQDTTLMLEKRSREYETRARIFKHELELRGIDIGLRDGNVONGHREDAVHAQ

NVAGHDKFTLMKSMNGTPROAQAFTDM"

/join(5475..5696,5827..6042,6389..6607,6684..6818,

6866..7000,7076..7378,7422..7571,7673..7760,8233..8503,

8615..8789,9895..9210)

/gene="F42A10.2"

## CDS

## CDS

## CDS

seq_documentation_block:	40852 bp	DNA	linear	INV 23-JAN-2002
LOCUS	CEFL1C1			
DEFINITION	Caenorhabditis elegans cosmid F11C1, complete sequence.			
ACCESSION	254270			
VERSION	254270.1	GI:3642282		
KEYWORDS	HTG; Denitification; Membrane glycoprotein; NIR; Nuclear hormone receptor.			
SOURCE	Caenorhabditis elegans.			
ORGANISM	Caenorhabditis elegans Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis. 1 (sites)			
REFERENCE	none.			
AUTHORS	Genome sequencing of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium			
TITLE	Science 282 (3596), 2012-2018 (1998)			
JOURNAL	99069613			
MEDLINE	The C.elegans Sequencing Consortium.			
REMARK	2 (bases 1 to 40852)			
REFERENCE	Palmer S.			
AUTHORS	Direct Submission			
TITLE	Submitted (29-SEP-1995) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: j.sanger@ac.uk or tw@nematode.wustl.edu			
JOURNAL	On Sep 22, 1998 this sequence version replaced gi:1001858.			
	Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information.			
	Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.			
	This sequence is the entire insert of clone F11C1. The start of this sequence (1..104) overlaps with the end of sequence AL02817. The end of this sequence (40749..40852) overlaps with the start of sequence AL008868.			
	For a graphical representation of this sequence and its analysis see: <a href="http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F11C1">http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F11C1</a>			
	IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.			
	Location/Qualifiers			
	1..40852			
	/organism="Caenorhabditis elegans"			
	/db_xref="taxon:6239"			
	/chromosome="X"			
	/clone="F11C1"			
	/join(980..1235,1367..1611)			
	/gene="F11C1.4"			
	/join(980..1235,1367..1611)			
	/gene="F11C1.4"			
	/codon_start=1			
	/protein_id="CAA91029.1"			
	/db_xref="GI:3875725"			
	/translation="MNISPLAYLVLPPLIIKFTHYVGNPDGDFYTFGRRLINLIAREGRRVQVFYFTLSLHNVLVLTPTSRCSSESKVVERSLADVOYKIDFVKEXLPGRNRYRMGHGDGAVMLSLTPYIKDQFNLRKAVCLPTEIRMTESHHGRRLKVVSVPR"			
	/join(4336..4463,5151..5312,5860..5935,5980..6062,6249..6327,6386..6513,6562..6717,6800..6929,7378..7692,7743..7816,7965..8229)			
	/gene="F11C1.3"			
	/join(4336..4463,5151..5312,5860..5935,5980..6062,6249..6327,6386..6513,6562..6717,6800..6929,7378..7692,7743..7816,7965..8229)			
	/gene="F11C1.3"			
	/note="contains similarity to Pfam domain: PF01130 (CD36			

```

family), Score=899.7, E-value=2.9e-267, N=1"
/codon_start=1
/protein_id="CAA91027.1"
/db_xref="GI:3875723"
/db_xref="SPTREMBL:O19344"
/translation="MRLPAIIIAAVLLVPAFGIFILLIPPEIAIPQSTIVDSQVYLQO
KSDQPCPGFYSKIPAVQITHTPILFNPDPENLYYGCAAPNNLEIGQVYTEREK
DPTQPRDNDIEFYVNNKINMYDPTSCCECKISDSVOFANTAMSGVITAGTLEFD
TGQDLITINGLILPMMKTEDTSDLRGCDGTIOKFGIOKTDVYQFOSLCKRNLH
HESKTVNSIPYTGKVEDDSIDAIKNGEIRFENEKVNYPNMGCLNHTKQNNKC
ALVDENQDNFCNACCDSSHVGTIVMPGWPQRCLEQWVPPREFALISAPHYGA
PVPYSEWVGILHDPDLAPRPFYINPATSTIGGFRRMLSTIVFQSLAMTMSNAP
NAMPSPFLITGVAMKDAVANTYIPNTATIPNIVLGLIGFTAAVLAVMLFCFYFR
SRNAKPPVLOOHRTEPTWSVAE"
complement(join(8545..8672,8990..9065,9345..9464,
9610..9678))
/gene="F11C1.2"
complement(join(8545..8672,8990..9065,9345..9464,
9610..9678))
/codon_start=1
/gene="F11C1.2"
/protein_id="CAA91026.1"
/db_xref="GI:3875722"
/db_xref="SPTREMBL:O19343"
/translation="MOVPHLSYGCADLMQIDPLNLANKKEEARMTLMERSVPEHI
TNGRATVPERERVFRITITPTSDSSNQNPTEEVKMKRMTVMGLVLENOI
VQKRREGLLALICAVITLAVSTLR"
complement(join(10039..10076,10909..11021,11077..11161,
11494..11602,11648..11683))
/gene="F11C1.1"
complement(join(10039..10076,10909..11021,11077..11161,
11494..11602,11648..11683))
/codon_start=1
/gene="F11C1.1"
/protein_id="CAA91025.1"
/db_xref="GI:3875721"
/db_xref="SPTREMBL:O19342"
/translation="MTESINSKSNQOYVMAKEEFPMDIDKSYFOKALTANPNVYP
DNNPGRPDVNFPCVCHLVASPCGEFREALSCOKTSSPEEMENACGEQLMA
PMECAMRCPTKNTSDNDEKSAEK"
complement(join(13974..14034,14082..14311,14643..14753,14801..14965,
15119..15286,15367..15483,15531..16099,16524..16761,
16817..17259,17304..17403,17556..17772,17820..18049,
18103..18220,18332..18684,18989..19141,19185..19272,
19317..19386,19431..19557,19607..19765,19816..19996,
20217..20322,20367..20612,20660..20879,20924..21025,
21071..21171,21222..21405,21489..21608,21660..21781,
21843..22047))
/gene="F11C1.5a"
complement(join(13974..14034,14082..14311,14643..14753,14801..14965,
15119..15286,15367..15483,15531..16099,16524..16761,
16817..17259,17304..17403,17556..17772,17820..18049,
18103..18220,18332..18684,18989..19141,19185..19272,
19317..19386,19431..19557,19607..19765,19816..19996,
20217..20322,20367..20612,20660..20879,20924..21025,
21071..21171,21222..21405,21489..21608,21660..21781,
21843..22047))
/gene="F11C1.5b"
complement(join(13974..14034,14082..14311,14643..14753,14801..14965,
15119..15286,15367..15483,15531..16099,16524..16761,
16817..17259,17304..17403,17556..17772,17820..18049,
18103..18220,18332..18684,18989..19141,19185..19272,
19317..19386,19431..19557,19607..19765,19816..19996,
20217..20322,20367..20612,20660..20879,20924..21025,
21071..21171,21222..21405,21489..21608,21660..21781,
21843..22047))
/notes="CDNA EST yk429a9.3 comes from this gene
CDNA EST yk529f9.3 comes from this gene"
/codon_start=1
/protein_id="CAA91031.1"
/db_xref="GI:3924721"
/db_xref="SPTREMBL:O19346"
/translation="MTRVSLIRLQIHAQRKLVLTASTSKRSKSTLTITSVVOLOAKN
PDVPAFYTGQNNPDDLPLHLRWLIDKDLKODVFLVLCVPEKIRLELVATLEATNR

```

```

EPEYLPITRDITADIKORREIRDGTAVYTDLCVBAALKGVLIVDVERARVLP
ILNNLEENEMOLDGGRFLMKHDKYDELCKISPFPELBPQATIPKNSFNLDLISV
IVYVNSEASTSRPRIGRAVYENTLTKINENPSYSAEITVPEYPAKVSLSDEKLTIN
TPTKREKSTHEPRRELKIDGALHTVQVPADEKRIOMKIPPSHTSPASVTPRAKO
ENIINDIALAESKGDFTLITGSGKRSITCEISRLNENYVTVLWIDLNAPTELOR
RHKKEGQWEDSITLTAARNDYCVIDYEWYHSYMSLSOLYTHREDLPQNR
LNEKRFILIERGVDLEARNERVOTEDSRVLYVESOSKEHNWENLSLIP
FIVPEPLMSLSDKSTISELTHSNTSEYDKLIRKVOITKNSDGLNRYTSLSLR
LITAKRDLMOGHEHRELVEKAAALSKLPQITKFEFHELTANFNFTITRKENEK
LDKIRANTIKEDALIPVNLPHNKKKHODELTEDMKRDKLSSLLILGOGGRKL
TDRELFILNRRQYMOHLRDTVOTLTQVLENGILIKEDSALVKARSGOLVILDE
ADKAPLHVAILKSLDGTGLVAGGRTRLRPASSFTEAKTDRVLRHNEFLIMLA
NRGAPFPGNNLFAVLELDLAIHMDHGRDSEBLAMIKRGPVPVKTILHQLTIEE
LRKRTDQGLQYPYSTRELVINVRHCNEFPDPLPEVCNRYVDSYSEDITATHEV
FOKGVPLGIVRTGESLISMERPLGEMKTIQKWSQKFEKSGFRTVAVEIERP
EKMEFGLKLDKOSNRNNVSELSHCVMKLPDESNTISDGIINDILVASVPELY
EKEFMDTESNKRIYFPRTHISDFOVDEINIKIIPPRAIKPRIRLAYGDNVILH
HEEANTFALISLPDKMAVYVTKOTSLGETLGAFTGSNSYMSLSMOOSHALLECG
REMLVYTDKRGVTVNGVTVPNIRKVPASAPDOOMTVSEMENTLLKVELGFTEK
INDLGGIDTVATRLDKNMTLVAEPPYFVQGNHGSVSGVLRKKSSEMTFHRPYSF
TDDRIRANDVSRNAPFGSTLVRAVAGFMPTECEVEKTK"
join(13974..14034,14082..14311,14643..14753,14801..14965,
15119..15286,15367..15483,15531..16099,16524..16761,
16817..17259,17304..17403,17556..17772,17820..18049,
18103..18220,18332..18684,18989..19141,19185..19272,
19317..19386,19431..19557,19607..19765,19816..19996,
20217..20322,20367..20612,20660..20879,20924..21025,
21071..21171,21222..21405,21489..21608,21660..21781,
21843..22047))
/gene="F11C1.5a"
/notes="similar to nlr like gene involved in
denitrification
CDNA EST yk12a1.3 comes from this gene
CDNA EST yk7e7.3 comes from this gene
CDNA EST yk7e7.5 comes from this gene
CDNA EST yk34c7.3 comes from this gene
CDNA EST yk12a1.5 comes from this gene
CDNA EST yk34c7.5 comes from this gene
CDNA EST yk154e5.3 comes from this gene
CDNA EST yk212d10.3 comes from this gene
CDNA EST yk212d10.5 comes from this gene
CDNA EST yk225b7.3 comes from this gene
CDNA EST yk225b7.5 comes from this gene
CDNA EST yk243b7.5 comes from this gene
CDNA EST yk348d4.5 comes from this gene
CDNA EST yk367e8.3 comes from this gene
CDNA EST yk367e8.5 comes from this gene
CDNA EST yk367e8.5 comes from this gene
CDNA EST yk420f3.3 comes from this gene
CDNA EST yk420f3.5 comes from this gene
CDNA EST yk529f9.5 comes from this gene"
/codon_start=1
/protein_id="CAA91030.1"
/db_xref="GI:3875726"
/db_xref="SPTREMBL:O19346"

```

```

alignment_scores:
  alignment: 8.00      Length: 8
  Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-08-973-363-14 x CEP11C1/rev ..
Align seg 1/1 to reverse of: CEP11C1 from: 1 to: 40852

35 GtutysargluThrltysglutlys 42
25767 GAAATAAGAGAAACAAAGAAATA 25744

seq_name: gb_htg:AC090028

```

seq\_documentation\_block: 50962 bp DNA linear HTG 16-OCT-2001  
LOCUS AC090028 Homo sapiens chromosome 3q clone RP11-29962, \*\*\* SEQUENCING IN  
DEFINITION  
ACCESSION AC090028  
VERSION AC090028.2 GI:14190614  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 50962)  
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Albrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbara,J.,  
Benton,J., Bimaga,K., Blankenburg,K., Bonin,D., Bouck,J.,  
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,  
Burck,P., Burckett,C., Burrell,K.L., Byrd,N.C., Carion,T.F.,  
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,  
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,  
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,  
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,  
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,  
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,  
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,  
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,  
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,  
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,  
Hollins,B., Homsli,F., Howard,S., Huber,J., Huliy,S., Hume,J.,  
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,  
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,  
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,  
Lousaged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,  
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S.,  
Ogun,M., Okunolu,G., Oragunde,N., Oviedo,R., Pace,A., Payton,B.,  
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,  
Quiles,M., Ren,Y., Rives,M., Rojas,A., RojudoKan,I., Ruiz,M.,  
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshari,N.,  
Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,  
Stone,H., Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K.,  
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,  
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Viallon,D., Vinson,R.,  
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,  
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,  
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 50962)  
Worley,K.C.  
Direct Submission  
Submitted (10-FEB-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On May 23, 2001 this sequence version replaced g1:12739654.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: HDEF  
Center clone name: RP11-29962  
----- Summary Statistics  
Sequencing vector: M13, L08821  
Chemistry: Dye-terminator Big Dye: 99% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 27482 bases at least Q40

Consensus quality: 35893 bases at least Q30  
Consensus quality: 39768 bases at least Q20  
Estimated insert size: 18665; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agorose-fp estimation  
Quality coverage: 0.2x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 48 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
647: contig of 647 bp in length  
648 747: gap of unknown length  
748 1436: contig of 689 bp in length  
1437 1536: gap of unknown length  
1537 2170: contig of 633 bp in length  
2170 2268: gap of unknown length  
2268 2928: contig of 659 bp in length  
2928 3028: gap of unknown length  
3028 3666: contig of 638 bp in length  
3666 3767: gap of unknown length  
3767 4441: contig of 675 bp in length  
4441 4541: gap of unknown length  
4541 5224: contig of 683 bp in length  
5224 5324: gap of unknown length  
5324 5953: contig of 629 bp in length  
5953 6053: gap of unknown length  
6053 6695: contig of 642 bp in length  
6695 6796: gap of unknown length  
6796 7431: contig of 636 bp in length  
7431 7531: gap of unknown length  
7531 8198: contig of 667 bp in length  
8198 8298: gap of unknown length  
8298 8949: contig of 651 bp in length  
8949 9049: gap of unknown length  
9049 9692: contig of 643 bp in length  
9692 9792: gap of unknown length  
9792 10508: contig of 716 bp in length  
10508 11694: gap of unknown length  
11694 11794: contig of 1086 bp in length  
11794 12537: gap of unknown length  
12537 12637: gap of unknown length  
12637 13282: contig of 645 bp in length  
13282 13382: gap of unknown length  
13382 14186: contig of 804 bp in length  
14186 14286: gap of unknown length  
14286 15068: contig of 782 bp in length  
15068 15168: gap of unknown length  
15168 16268: contig of 1100 bp in length  
16268 17150: gap of unknown length  
17150 17250: contig of 782 bp in length  
17250 18072: gap of unknown length  
18072 18172: contig of 822 bp in length  
18172 18668: gap of unknown length  
18668 18968: contig of 696 bp in length  
18968 19868: gap of unknown length  
19868 19968: contig of 900 bp in length  
19968 20865: gap of unknown length  
20865 20965: contig of 897 bp in length  
20965 21680: gap of unknown length  
21680 21780: contig of 715 bp in length  
21780 22809: gap of unknown length  
22809 22910: contig of 1029 bp in length  
22910 24084: gap of unknown length  
24084 24184: contig of 1175 bp in length  
24184 25112: gap of unknown length  
25112: contig of 928 bp in length





```
repeat_region 5046..5435
/note="LMC2 repeat: matches 5910..6324 of consensus"
repeat_region 5439..5813
/note="MLT1B repeat: matches 1..390 of consensus"
repeat_region 5828..6022
/note="MST-INTERNAL repeat: matches 1443..1637 of consensus"
repeat_region 6023..6314
/note="AluX repeat: matches 1..293 of consensus"
repeat_region 6311..6391
/note="MLT1-INTERNAL repeat: matches 1046..1111 of consensus"
repeat_region 6392..6685
/note="AluSx repeat: matches 1..290 of consensus"
repeat_region 6686..6788
/note="MLT1-INTERNAL repeat: matches 961..1046 of consensus"
repeat_region 6790..7091
/note="AluSp repeat: matches 1..299 of consensus"
repeat_region 7539..8133
/note="MLT1B-internal repeat: matches 1..728 of consensus"
repeat_region 8168..8363
/note="MLT1B repeat: matches 194..390 of consensus"
repeat_region 8427..8451
/note="MLT1B repeat: matches 168..188 of consensus"
repeat_region 8452..8758
/note="AluJo repeat: matches 1..301 of consensus"
repeat_region 8759..8910
/note="MLT1B repeat: matches 1..168 of consensus"
repeat_region 8911..8998
/note="LMC2 repeat: matches 5826..5911 of consensus"
repeat_region 8999..9299
/note="AluSx repeat: matches 1..300 of consensus"
repeat_region 9300..9949
/note="LMC2 repeat: matches 5189..5826 of consensus"
repeat_region 9950..10223
/note="AluJo repeat: matches 1..298 of consensus"
repeat_region 10224..10633
/note="LMC2 repeat: matches 4787..5189 of consensus"
repeat_region 10634..10920
/note="AluSx repeat: matches 12..296 of consensus"
repeat_region 10921..11272
/note="LMC2 repeat: matches 4362..4787 of consensus"
repeat_region 11292..11412
/note="FLAM_A repeat: matches 13..128 of consensus"
repeat_region 11413..12017
/note="LMC2 repeat: matches 3787..4359 of consensus"
repeat_region 12046..12343
/note="AluSg1 repeat: matches 1..294 of consensus"
repeat_region 12582..18699
/note="LIPAs repeat: matches 11..6143 of consensus"
repeat_region 18731..19020
/note="Aluub repeat: matches 1..293 of consensus"
misc_feature 18754..19116
/note="match: GSS: Em:AQ521768"
repeat_region 19702..20069
/note="THE1B repeat: matches 1..364 of consensus"
repeat_region 20223..20631
/note="MLT2FB repeat: matches 1..410 of consensus"
repeat_region 21092..21183
/note="MIR repeat: matches 23..112 of consensus"
repeat_region 21243..21463
/note="MER20 repeat: matches 2..195 of consensus"
repeat_region 22606..22805
/note="MIR repeat: matches 49..250 of consensus"
repeat_region 22909..23196
/note="AluJo repeat: matches 1..294 of consensus"
repeat_region 23505..23562
/note="29 copies 2 mer at 70% conserved"
repeat_region 23700..23827
/note="LIME repeat: matches 5678..5804 of consensus"
repeat_region 23841..24023
/note="L2 repeat: matches 1904..2095 of consensus"
```

---

```
repeat_region 24642..24946
/note="AluSx repeat: matches 1..300 of consensus"
repeat_region 25088..25297
/note="MIR repeat: matches 32..247 of consensus"
repeat_region 27241..27530
/note="AluJo repeat: matches 1..288 of consensus"
repeat_region 28031..28100
/note="L2 repeat: matches 2625..2693 of consensus"
repeat_region 28236..28482
/note="MER8 repeat: matches 4..238 of consensus"
repeat_region 28635..28670
/note="18 copies 2 mer lt 80% conserved"
repeat_region 29295..29412
/note="59 copies 2 mer cc 65% conserved"
misc_feature complement(29374..29836)
/note="match: GSS: Em:AQ493139"
repeat_region 30071..30203
/note="LMC5 repeat: matches 7762..7893 of consensus"
misc_feature 31005..31642
/note="match: GSS: Em:B6029"
misc_feature 31005..31469
/note="match: GSS: Em:AQ232280"
repeat_region 32407..32692
/note="Aluub repeat: matches 1..300 of consensus"
repeat_region 32729..33022
/note="AluSx repeat: matches 4..300 of consensus"
repeat_region 33097..33203
/note="MLT1D repeat: matches 396..505 of consensus"
repeat_region 33204..33510
/note="AluY repeat: matches 1..305 of consensus"
repeat_region 33511..33886
/note="MLT1D repeat: matches 40..396 of consensus"
repeat_region 33990..34197
/note="MIR repeat: matches 13..252 of consensus"
repeat_region 35679..35938
/note="MIR repeat: matches 8..262 of consensus"
repeat_region 36923..37001
/note="AluSc repeat: matches 223..299 of consensus"
repeat_region 37212..37407
/note="MER5A repeat: matches 1..188 of consensus"
repeat_region 38378..38435
/note="29 copies 2 mer cc 70% conserved"
repeat_region 38441..38712
/note="AluJo repeat: matches 1..280 of consensus"
misc_feature complement(38574..39022)
/note="match: GSS: Em:AQ572560"
```

alignment\_scores:

Quality:	8.00	Length:	8
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment\_block:

US-08-973-363-14 x HSDJ223E3/rev ..

Align seg 1/1 to reverse of: HSDJ223E3 from: 1 to: 59621


```
36 LysArgGluThrLysGluLysGlu 43
|||||
54121 AAGAGAGAAACAAAGAGAAAGAA 54098
```

Mon Aug 5 11:51:44 2002

us-08-973-363-14.olip2n.rge

---

Page 17











```
Align seg 1/1 to: AAS1339 from: 1 to: 657
      35 GlutylsATGtUThrLySGlulys 42
      ||||||||||||||||||
496 GAAAAAGCAACGAAAGAGAGAA 519

seq_name: /STD1/gcgdata/hold-geneseq/geneseq-nembl/NA2001A.DAT.AAK89386
seq_documentation_block:
ID   AAK89386 standard; DNA; 2121 BP.
XX
AC   AAK89386;
XX
DT   05-NOV-2001 (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 2962.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   MO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001MO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205513.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226688.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228824.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
```

```
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   27-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   01-NOV-2000; 2000US-0244647.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246539.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
```



PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250130.  
 PR 05-DEC-2000; 2000US-0251088.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI, 2001-502630/55.  
 XX  
 PT Polynucleotides encoding digestive system antigens, useful for  
 PT diagnosing, treating, preventing and/or prognostic disorders of the  
 PT digestive system, particularly cancer and cancer metastases -  
 PT  
 PS Disclosure; SEQ ID NO 2962; 986bp; English.  
 PS  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human digestive system antigens. These can be used in the  
 CC diagnosis, treatment and prevention of digestive system disorders,  
 CC including cancer. Meckel's diverticulum, bacterial or parasitic  
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
 CC ulcerative colitis.. The present sequence is a genomic DNA fragment  
 CC encoding a digestive system antigen of the invention.  
 XX  
 XX Sequence 2121 BP; 906 A; 306 C; 407 G; 502 T; 0 other;  
 S0  
 Alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-14 x AAK89386 ..  
 Align seg 1/1 to: AAK89386 from: 1 to: 2121  
 35 GIULYSARGLunThrLySGluLys 42  
 ||||||||||||||||||||  
 1368 GAAAAAAGCGAAGCGAAGGAAAA 1391  
 seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: AAC25842  
 seq\_documentation\_block:  
 ID AAC25842 standard; CDNA; 294 BP.  
 AC AAC25842;  
 XX  
 XX . 06-OCT-2000 (first entry)  
 DE Human secreted protein 5' EST. SEQ ID NO: 29917.  
 XX  
 XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 XX gene therapy; chromosome mapping; ss.  
 XX  
 XX Homo sapiens.

```

PN      EPI03401-A2.
PD
XX      06-SEP-2000.
XX
PF      21-FEB-2000; 2000EP-0200610.
XX
PE      26-FEB-1999; 99US-0122487.
XX
PR      (GEST ) GENSET.
XX
PA      Dumas Milne Edwards J, Duclert A, Giordano J;
PI      WPI; 2000-500381/45.
XX
XX      New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT      obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PR      diagnostic, forensic, gene therapy and chromosome mapping procedures -
PT      Claim 1; SEQ ID 29917; 71np + CD-ROM; English.
XX
PS      The present sequence is one of a large number of 5' ESTs derived from
CC      mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC      identified within the present sequence. The 5' ESTs were prepared from
CC      total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC      sequences usually correspond mainly to the 3' untranslated region (UTR)
CC      of the mRNA because they are often obtained from oligo-dT primed cDNA
CC      libraries. Such ESTs are not well suited for isolating cDNA sequences
CC      derived from the 5' ends of mRNAs and even in those cases where longer
CC      cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC      5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC      used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC      in diagnostic, forensic, gene therapy and chromosome mapping procedures
CC      They are used to obtain upstream regulatory sequences and to design
CC      expression and secretion vectors.
XX
SQ      Sequence 294 BP; 96 A; 51 C; 72 G; 75 T; 0 other.
XX

Alignment_scores:
                Quality:   7.00          Length:       7
                Ratio:     1.000         Gaps:           0
Percent Similarity: 100.000    Percent Identity: 100.000

Alignment_block:
US-08-973-363-14 x AAC25842 ..

Align seg 1/1 to: AAC25842 from: 1 to: 294

29 ILeYtSYsgIuaIagIyGlu 35
|||||
170 ATTAAGAAACAGACTGTGCAG 190

seq_name: /SIBSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT.AAK70025
seq_documentation_block:
ID ID AAK70025 standard; DNA; 312 BP.
AC
XX      AAK70025;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24837.
KW      Human; Immune; haematopoietic; Immune/haematopoietic antigen; cancer;
XX      cytostatic; gene Therapy; vaccine; metastasis; ds.
XX
OS      Homo sapIens.
XX      WO200157182-A2.
XX      PN
XX      09-AUG-2001.
XX

```

PE 17-JAN-2001; 2001MO-US01354.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 29-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249267.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC;  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI, 2001-483426/52.  
DR  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT

PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
PS Disclosure: SEQ ID NO 24837; 3071pp + Sequence Listing: English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82191 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 312 BP; 133 A; 40 C; 82 G; 57 T; 0 other;

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x AAK70025 ..  
Align seg 1/1 to: AAK70025 from: 1 to: 312

37 ArgGluThrLysGluLysGlu 43  
|||||  
50 AGGAGAGCCAGCAGAGAGAGAG 70

seq\_name: /SIDSL/gcgcdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT: AAS39061  
seq\_documentation\_block:  
ID AAS39061 standard; cDNA; 350 BP.  
XX  
AC AAS39061;  
XX  
DE 17-DEC-2001 (first entry)  
XX  
DE Novel human diagnostic and therapeutic gene #2119.  
XX  
KW Human; cancer; breast; lung; colon; prostate; cytostatic; diagnostic; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20016753-A2.  
XX  
PD 13-SEP-2001.  
XX  
PE 09-MAR-2001; 2001WO-US07787.  
XX  
PR 09-MAR-2000; 2000US-0188609.  
XX  
PA (CHIR ) CHIRON CORP.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Rainard C, Randazzo E, Kennedy GC, Pot D, Kassam A, Lamson G;  
PI Dimeac R, Chtvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leshnowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;  
XX  
XX WPI; 2001-530177/58.

PT New polynucleotides and polypeptides, useful for diagnosis and  
PT treatment of breast, lung and colon cancer -  
PS Claim 1; Page 1126; 1193pp; English.  
XX  
XX  
CC The invention relates to new polynucleotides and polypeptides, useful for  
CC diagnosis and treatment of breast, lung and colon cancer. The sequences  
CC can be used in detecting differentially expressed genes correlated with a  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample derived from a  
CC cell suspected of being cancerous. They can also be used to inhibit  
CC tumor growth by modulating expression of a gene product. AAS3943-  
CC AAS39338 represent novel human diagnostic and therapeutic coding  
CC sequences of the invention.  
XX  
SQ Sequence 350 BP; 53 A; 84 C; 60 G; 153 T; 0 other;

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x AAS39061/rev ..  
Align seg 1/1 to reverse of: AAS39061 from: 1 to: 350

31 LysGluAlaGlyGluLysArg 37  
|||||  
260 AAGAGAGCAGCGAGAGAGGA 240

seq\_name: /SIDSL/gcgcdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: AAF10282  
seq\_documentation\_block:  
ID AAF10282 standard; cDNA; 427 BP.  
XX  
AC AAF10282;  
XX  
DE 13-MAR-2001 (first entry)  
XX  
DE Fusarium venenatum EST SEQ ID NO:2805.  
XX  
KW Multiple gene expression; filamentous fungal cell; EST.  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catalytic pathway engineering; ss.  
XX  
OS Fusarium venenatum.  
XX  
PN WO200056762-A2.  
XX  
PD 28-SEP-2000.  
XX  
PE 22-MAR-2000; 2000WO-US07781.  
XX  
PR 22-MAR-1999; 99US-0273623.  
XX  
PA (NOVO ) NOVO NORDISK BIOTECH INC.  
XX  
PA (NOVO ) NOVO NORDISK AS.  
XX  
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;  
XX  
XX WPI; 2000-594572/56.

Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags -  
PS Claim 86; Page 1400; 3161pp; English.  
XX  
XX The present invention describes a method for monitoring differential

CC expression of genes in a first filamentous fungal (FF) cell relative to  
CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring  
CC the global expression of genes from FF cells allows the production  
CC potential of the microorganisms to be improved. New genes may be  
CC discovered, possible functions of unknown open reading frames can be  
CC identified and gene copy number variation and stability can be  
CC monitored. The expression of genes can be used to study how FF cells  
CC adapt to changes in culture conditions, environmental stress, spore  
CC morphogenesis, recombination, metabolic or catabolic pathway  
CC engineering. Using ESTs provides several advantages over genomic or  
CC random cDNA clones including elimination of redundancy as one spot on an  
CC array equals one gene or open reading frame, and organisation of the  
CC microarrays based on function of the gene products to facilitate  
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
CC all specifically claimed in the present invention.

SQ Sequence 427 BP; 120 A; 76 C; 97 G; 134 T; 0 other;

## alignment\_scores:

Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-14 x AAF10282/rev ..

Align seg 1/1 to reverse of: AAF10282 from: 1 to: 427

39 Thrltysglutylsgluasnlys 45  
|||||||  
165 ACAAAAGAAAAAGAAACAA 145

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: AAC00952

## seq\_documentation\_block:

ID AAC00952 standard; cDNA; 468 BP.

AC AAC00952;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 950.

KW Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR P-PSDB; AAG00946.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
PS Claim 1; SEQ ID 950; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.

SQ Sequence 468 BP; 131 A; 89 C; 116 G; 131 T; 1 other;

## alignment\_scores:

Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-14 x AAC00952 ..

Align seg 1/1 to: AAC00952 from: 1 to: 468

13 Illelysalglutylsgluasn 19  
|||||||  
357 ATCAAGCTGAAAGAAAT 377

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT: AAC00953

## seq\_documentation\_block:

ID AAC00953 standard; cDNA; 542 BP.

AC AAC00953;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 951.

KW Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR P-PSDB; AAG00947.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PT Claim 1; SEQ ID 951; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or poly(A<sup>+</sup>) RNAs  
CC derived from 33 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC cDNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.

SQ Sequence 542 BP: 151 A; 105 C; 130 G; 155 T; 1 other;

Alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

Alignment\_block:  
US-08-973-363-14 x AAC00953 ..

Align seg 1/1 to: AAC00953 from: 1 to: 542

13 TlclysAlaGluLysGluAsn 19  
|||||  
431 ATCAGGCTGAAAAAGAAAT 451

seq\_name: /SIDSI/gcdata/hold-geneseq/genesequ-emb1/NA2000.DAT:AAC54696

seq\_documentation\_block:  
ID AAC54696 standard; DNA; 695 BP.

XX AAC54696;  
AC  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 78761.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
OS  
PN EP1033405-A2.  
PN  
PD 06-SEP-2000.  
PD  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
PF  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128645.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132488.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134220.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137529.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139753.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139829.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 10-AUG-1999; 99US-0147935.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 23-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151348.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158923.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160981.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:  
Quality: 7.00  
Ratio: 1.000  
Percent Similarity: 100.000  
Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x AAC54696 ..  
Align seg 1/1 to: AAC54696 from: 1 to: 695

26 GIUUEGLYILELYSLYGLU 32  
|||||  
190 GAGATTGGATCAAAAGAG 210

seq\_name: /SIDSI/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAC52409  
seq\_documentation\_block:  
ID AAC52409 standard; DNA: 696 BP.  
XX  
AC AAC52409;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 71249.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 09-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147933.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151080.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151920.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 14-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.

PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x AAC52409 ..

Align seg 1/1 to: AAC52409 from: 1 to: 696

26 Gluileglylilelysglu 32  
|||||  
189 GAGATTGGGATCAAAAGAG 209

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:AAF11881

seq\_documentation\_block:  
ID AAF11881 standard; cDNA; 710 BP.  
XX AAF11881;  
XX  
DT 13-MAR-2001 (first entry)  
XX  
DE Aspergillus oryzae EST SEQ ID NO:4404.  
XX  
KW Multiple gene expression; filamentous fungal cell; EST;  
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;  
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;  
KW culture condition; environmental stress; spore morphogenesis;  
KW metabolic pathway engineering; catabolic pathway engineering; ss.  
XX  
OS Aspergillus oryzae.  
XX  
PN WO200056762-A2.  
XX  
PD 28-SEP-2000.  
XX  
PF 22-MAR-2000; 2000WO-US07781.  
XX  
PR 22-MAR-1999; 99US-0273623.  
XX  
PA (NOVO ) NOVO NORDISK BIOTECH INC.  
(NOVO ) NOVO NORDISK AS.  
XX  
PI Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;  
XX  
DR WPI; 2000-594572/56.  
XX  
PT Monitoring differential expression of genes in filamentous fungal cells  
PT uses fluorescence-labeled nucleic acids isolated from the cells and a  
PT substrate of expressed sequence tags -  
XX  
PS Claim 88; Page 1883-1884; 3161pp; English.  
XX  
CC The present invention describes a method for monitoring differential  
CC expression of genes in a first filamentous fungal (FF) cell relative to

CC expression of the same genes in one or more second filamentous fungal  
CC cells. The method uses fluorescence-labeled nucleic acids isolated from  
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs  
CC are used in the methods for monitoring differential expression of genes  
CC in a first filamentous fungal (FF) cell relative to expression of the  
CC same genes in one or more second filamentous fungal cells. Monitoring  
CC the global expression of genes from FF cells allows the production  
CC potential of the microorganisms to be improved. New genes may be  
CC discovered, possible functions of unknown open reading frames can be  
CC identified and gene copy number variation and stability can be  
CC monitored. The expression of genes can be used to study how FF cells  
CC adapt to changes in culture conditions, environmental stress, spore  
CC morphogenesis, recombination, metabolic or catabolic pathway  
CC engineering. Using ESTs provides several advantages over genomic or  
CC random cDNA clones including elimination of redundancy as one spot on an  
CC array equals one gene or open reading frame, and organisation of the  
CC microarrays based on function of the gene products to facilitate  
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from  
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus  
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and  
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are  
CC all specifically claimed in the present invention.  
XX  
SQ Sequence 710 BP; 135 A; 195 C; 148 G; 231 T; 1 other;

alignment\_scores:  
Quality: 7.00 Length: 7  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x AAF11881/rev ..

Align seg 1/1 to reverse of: AAF11881 from: 1 to: 710

31 LysGluAlaGlyGluLysArg 37  
|||||  
233 AAAAGAGCGGTCGAAAAGAG 213

seq\_name: /SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAI97871

seq\_documentation\_block:  
ID AAI97871 standard; cDNA; 721 BP.  
XX AAI97871;  
XX  
DT 13-NOV-2001 (first entry)  
XX  
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 3946.  
XX  
KW Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20016719-A1.  
XX  
PD 13-SEP-2001.  
XX  
PF 02-MAR-2001; 2001WO-JP01629.  
XX  
PR 07-MAR-2000; 2000JP-0159195.  
XX  
PA (CHIB-) CHIBA PREFECTURE.  
(HISM ) HISAMITSU PHARM CO LTD.  
XX  
PI Nakagawara A;  
XX  
DR WPI; 2001-565584/63.  
XX  
PT Nucleic acids originating in gene expressed in human neuroblastoma,  
PT useful as probe or primer in diagnosing prognosis of human  
PT neuroblastoma, malignancy and susceptibility indicator or tumour marker



PT for anti-cancer agents -  
 XX  
 PS Claim 1; Page 2902; 2979pp; Japanese.  
 XX  
 CC The invention relates to novel genes (AA193926-AA197863) expressed in  
 CC human neuroblastoma. The nucleic acids are applicable as a probe or  
 CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
 CC susceptibility indicators or tumour markers for anti-cancer agents. The  
 CC gene information for diagnosing prognosis is related to factors similar  
 CC to that for N-myc and trkA genes.  
 XX  
 SQ Sequence 721 BP; 151 A; 200 C; 176 G; 190 T; 4 other;  
 alignment\_scores:  
 Quality: 7.00 Length: 7  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-14 x AA197871/rev ..  
 Align seg 1/1 to reverse of: AA197871 from: 1 to: 721  
 32 GUAAGAGLGLTYSARGLU 38  
 ||||||||||||||||||  
 434 GAGGCTGGGCAAGACGAGNA 414  
 seq\_name: /SIDS1/gcdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:AAH77229  
 seq\_documentation\_block:  
 ID AAH77229 standard; DNA; 729 BP.  
 XX  
 AC AAH77229;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE DNA of atRSP41 5'-UTR.  
 XX  
 KW atRSP41 5'-UTR; thale cress; transgenic plant; virus resistance; ds;  
 KW altered reproductive capability; insect resistance; disease resistance;  
 KW Arabidopsis thaliana arginine/serine-type plant splicing factor; atRSP;  
 KW herbicide tolerance; antibiotic resistance.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT 5'UTR 1..729  
 FT Intron 83..724  
 FT /tag= b  
 FT /number= "1"  
 FT  
 XX  
 MO200175129-A1.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 02-APR-2001; 2001MO-BP03735.  
 XX  
 PR 03-APR-2000; 2000GB-0008120.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Barta A, Lopato S, Kalyna M:  
 XX  
 DR MPI; 2002-010793/01.  
 XX  
 PT DNA sequences comprising atRSP gene promoters, useful for expressing a  
 PT nucleotide sequence of interest for conferring a desirable phenotypic  
 PT trait to a plant, e.g. virus, insect or disease resistance, are new.  
 XX  
 PS Claim 5; Page 30; 34pp; English.  
 XX

CC This polynucleotide sequence represents the DNA of the atRSP41 5'-UTR.  
 CC The invention relates to novel DNA sequences that function as promoters  
 CC of transcription of associated nucleotide sequences in plants. The  
 CC invention provides a DNA sequence capable of driving expression of an  
 CC associated nucleotide sequence, wherein said DNA sequence is obtainable  
 CC from genes of the atRSP (Arabidopsis thaliana arginine/serine-type plant  
 CC splicing factor) gene family. The invention further provides recombinant  
 CC DNA molecules comprising an atRSP41 promoter, the invention also provides  
 CC expression vectors and transgenic plants comprising the promoters of the  
 CC invention. The DNA sequence of the invention is useful for expressing a  
 CC nucleotide sequence of interest. The nucleotide sequences encode proteins  
 CC for conferring a desirable phenotypic trait to a plant transformed with  
 CC the protein, where such traits include antibiotic resistance, virus  
 CC resistance, insect resistance, disease resistance, or resistance to other  
 CC pests, herbicide tolerance, improved nutritional value, improved  
 CC performance in an industrial process or altered reproductive capability.  
 XX  
 SQ Sequence 729 BP; 168 A; 90 C; 137 G; 334 T; 0 other;  
 alignment\_scores:  
 Quality: 7.00 Length: 7  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-08-973-363-14 x AAH77229/rev ..  
 Align seg 1/1 to reverse of: AAH77229 from: 1 to: 729  
 35 GLUYSARGGLUHTLYSGLU 41  
 ||||||||||||||||||  
 145 GAGAAGAGACGACCAAGAC 125





Genetically and physically anchored EST resources for  
Barley genomics. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/gppages/bgn/31/cover.html)"

BASE COUNT 257 a 216 c 345 g 93 t

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x BG309136 ..

Align seg 1/1 to: BG309136 from: 1 to: 911

34 GtlylgsArgGluThrLysGluLys 42  
|||||  
827 GCGGAAAAGAGAGAAAGAGAGAG 853

seq\_name: gb\_est2:BG283942

seq\_documentation\_block:

LOCUS BG283942 1227 bp mRNA linear EST 21-FEB-2001  
DEFINITION 602408112F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4520114 5',  
mRNA sequence.

ACCESSION BG283942  
VERSION BG283942.1 GI:13034392

KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1227)

AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTF  
CDNA library Preparation: Life Technologies, Inc.  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLAM10417 row: h column: 03  
High quality sequence stop: 168.  
Location/Qualifiers  
1. 1227

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:4520114"  
/clone\_lib="NIH\_MGC\_91"  
/tissue\_type="adenoecarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Prostate; Vector: pCMV-SPORT6; Site\_1: NotI;  
Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.4 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH\_MGC library."

BASE COUNT 587 a 208 c 388 g 44 t

ORIGIN

alignment\_scores:  
Quality: 9.00 Length: 9  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x BG283942 ..

Align seg 1/1 to: BG283942 from: 1 to: 1227

34 GtlylgsArgGluThrLysGluLys 42  
|||||  
629 GCGGAAAAGAGAGAGAGAGAGAG 655

seq\_name: gb\_est2:BI010841

seq\_documentation\_block:

LOCUS BI010841 198 bp mRNA linear EST 13-JUN-2001  
DEFINITION MR2-EN0092-110101-006-d03\_1 EN0092 Homo sapiens cDNA, mRNA  
sequence.

ACCESSION BI010841  
VERSION BI010841.1 GI:14414912

KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 198)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR2&ct=MR2-EN0092-  
110101-006-d03\_1&f3=2001-01-11&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 56.  
Location/Qualifiers  
1. 198

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="EN0092"  
/dev\_stage="Adult"  
/note="Organ: Lung,normal; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESRES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 38 a 51 c 31 g 78 t

ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-08-973-363-14 x BI010841/rev ..

Align seg 1/1 to reverse of: BI010841 from: 1 to: 198

35 GtlylgsArgGluThrLysGluLys 42

```

seq_name: gb_est2:BE787833
seq.documentation_block:
LOCUS BE787833 251 bp mRNA EST 20-OCT-2000
DEFINITION B01482830F1 NIH_MGC_68 Homo sapiens CDNA clone IMAGE:3885463 5',
ACCESSION mRNA sequence.
VERSION BE787833
KEYWORDS BE787833.1 GI:10209031
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.temail.nih.gov
Tissue Procurement: DCD/DTP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LNLN)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM9660 row: 1 column: 08
High quality sequence stop: 188.
Location/Qualifiers
1..251
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3885463"
/clone_1lb="NIH-MGC_68"
/tissue_type="large_cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung Vector: pCMV-Sport6, Site:1; Notf:
Site:2; Salt: Cloned unidirectionally. Primer: oligo GT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 93 a 58 c 74 g 26 t
ORIGIN
Alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-973-363-14 x BE787833 ..
Align seg 1/1 to: BE787833 from: 1 to: 251
16 GtutysGtutysAngtutysasp 23
|||||
213 GAAAAAGCAATGAGAAAGAC 26
seq_name: gb_est1:BB416876
seq.documentation_block:
LOCUS BB416876 264 bp mRNA EST 16-JUL-2000
DEFINITION BB416876 RIKEN full-length enriched, 7 days embryo MUS musculus
CDNA clone C430048407 3', mRNA sequence.
ACCESSION BB416876
VERSION BB416876.1 GI:9238231
KEYWORDS EST.
SOURCE house mouse.
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE	Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus
AUTHORS	1 (Pages 1 to 264) Komono, H., Aizawa, K., Akhizari, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hata, A., Hayatsu, N., Hirose, T., Horie, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kituchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Saki, C., Saito, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomiunga, N., Toya, T., Tsumoda, Y., Watanabe, K., Watanabe, S., Yamamura, T., Yamazaki, I., Yanai, R., Yasunishi, A., Yokota, T., Yoshida, R., Yoshiaki, A., Yoshino, M., Yamatsuta, M., and Hayashizaki, Y.
TITLE	RIKEN Mouse ESTs (Komono, H., et al.)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Toshinhide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan Tel.: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/ Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagakawa, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Thermosensitization and thermostabilization of chemolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh, M., Kitsuai, T., Akiyama, Y., Shibata, K., Iizawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site ( <a href="http://genome.ritc.riken.go.jp">http://genome.ritc.riken.go.jp</a> ) for further details
FEATURES	Location/Qualifiers 1..264 /organism="Mus musculus" /db_xref="taxon:10090" /clone_id="C430048M07" /clone_1lb="RIKEN full-length enriched, 7 days embryo" /dev_stage="7 days embryo" /dev_host="DH10B" /note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGACAGACATCCACAGCCTCTTTTCTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGACAGCTTCGACGCTTAATTAATTAATCCGCCGCCGCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a pBluescript KS(+) after bulk excision from Lambda FIC 1."
BASE COUNT	45 a 87 c 36 g 96 t
ORIGIN	
alignment_scores:	Quality: 8.00 Length: 8
	Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000	Percent Identity: 100.000
alignment_block:	

alignment\_block:

31 LysGluAlaGlyGluLysArgGlu 38  
 |||||  
 15 AAGGAGCGAGGGAGAGAGCGAGAG 38

```

seq_name: gb_est2:D71396

seq_documentation_block:
  LOCUS D71396 300 bp mRNA EST 18-OCT-1998
  DEFINITION CEMK0747R yuji kohara unpublished cDNA:Strain N2 hermaphrodite
  ACCESSION D71396
  VERSION D71396.1 GI:1111103
  KEYWORDS EST.
  ORGANISM Caenorhabditis elegans.
  Eukaryota; Metazoa; Chordata; Chromadorea; Rhabdilita; Rhabditoidea;
  Rhabdilitidae; Pelodidae; Caenorhabditis.
  1 (bases 1 to 300)
  Kohara,T., Mitsuki,H., Nishigaki,A., Mochishashi,T., Sugimoto,A. and
  Tabara,H.
  TITLE Toward an expression map of the C. elegans genome
  JOURNAL Unpublished (1994)
  COMMENT Contact: Yuji Kohara
  Genome Biology Lab.
  National Institute of Genetics
  Yata 1111, Mishima, Shizuoka 411, Japan
  Tel: 81-559-81-6854
  Fax: 81-559-81-6855
  Email: ykohara@lab.nig.ac.jp.
  Location/Qualifiers
  1..300
  /organism="Caenorhabditis elegans"
  /strain="N2"
  /db_xref="taxon:6239"
  /clone="YK74f7"
  /clone_1b="Yuji Kohara unpublished cDNA:Strain N2
  hermaphrodite embryo"
  /sex="hermaphrodite"
  /dev_stage="embryo"

BASE COUNT 122 a 45 c 67 g 58 t 8 others
ORIGIN

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-08-973-563-14 x D71396 ..

Align seg 1/1 to: D71396 from: 1 to: 300

35 GULYsARGGLuThrYsGLuYs 42
|||||
20 GAAAAAACGCAACGAAAGAAAA 43

seq_name: gb_est1:AW507143

seq_documentation_block:
  LOCUS AW507143 303 bp mRNA EST 01-JUN-2000
  DEFINITION EST00470 Plasmid subtractive library of Rat Cerebrum (stroke)
  ACCESSION AW507143
  VERSION AW507143.1 GI:8133308
  KEYWORDS EST.
  SOURCE Norway rat
  ORGANISM Rattus norvegicus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
  Rattus.
  1 (bases 1 to 303)
  Zhao,B., Wang,X.M., Zhang,O., Zhu,Z.M., Zhu,S.J. and Hui,R.T.
  TITLE Cloning of genes responsible for Stroke
  JOURNAL Unpublished (2000)
  COMMENT Contact: Bin Zhao
  Molecular Medical Center for Cardiovascular Disease

```

**FEATURES**

**SOURCE**

POLYA-No.

Location/Qualifiers

1. 303

/organism="Rattus norvegicus"

/strain="Wistar"

/db\_xref=taxon:10116

/clone="3rdZ"

/clone\_1b="Plasmid Subtractive Library of Rat Cerebrum (stroke)"

/sex="male"

/tissue\_type="Cerebrum"

/dev\_stage="8-12 weeks old"

/note=Vector: pMD-T (Takara, D504 CA): The rat stroke model was developed by Wang Xianmei, Zhu Zhiming, Zhu Shanjun et al at 3rd Military Medical University Chongqing , China. The Suppression Subtraction Hybridization (SSH) was conducted at Fu Wai Hospital by Zhao Bin and Wang Xianmei in Beijing, China. Message RNA was extracted from pooled whole cerebrum from stroke and control rats respectively. SSH had been conducted following manufacture's manual using 5ug mRNA. The SSH library was constructed by Zhao Bin and Wang Xianmei using pMD18-T vector. Randomly picked clones were sequenced at ABI 377 using M13 primer TACGGATTAATTTCACAGG) by Wang Xianmei.

BASE COUNT      90 a    90 c    90 g    90 t

ORIGIN

Alignment\_scores:

Quality: 8.00      Length: 8

Ratio: 1.000      Gaps: 0

Percent Similarity: 100.000      Percent Identity: 100.000

Alignment\_block:

US-08-973-363-14 x AW507143 ..

Align seq 1/1 to: AW507143 from: 1 to: 303

11 LysLysIleLysAlaGluGln 18  
|||||  
|||

seq\_name: gb\_est2:B604006

seq\_documentation\_block:

LOCUS B604006 315 bp mRNA linear EST 14-AUG-2001

DEFINITION ESF503096 plasmodium yoelii sporozoite cDNA Plasmodium yoelii cDNA

ACCESSION B604006

VERSION B604006.1 GI:15154020

KEYWORDS EST.

SOURCE

ORGANISM

Eukaryota; Alveolata; Apicomplexa; Hemosporida; Plasmidium.

1 (bases 1 to 315)

Ribeiro,S.H.I., Gardner,M.J., Brown,S.M., Ross,J., Matuschewski,K., Holman,J.V., Adams,D.H., Quackenbush,J., Cho,J., Carucci,D.J., Hellebrand,N., and Nussenzweig,V.

Exploiting the transcriptome of the malaria sporozoite stage

Proc. Natl. Acad. Sci. U. S. A. 98 (17), 9895-9900 (2001)

21396655

Contact: Malcolm J. Gardner

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301 838 3519

Fax: 301 838 0208  
Email: gardner@tigr.org  
Request for clones, please contact: Stefan Kappe,  
kappes01@popmail.med.nyu.edu Michael Heidelberger Division,  
Department of Pathology New York University School of Medicine.

## FEATURES

## source

1. .315  
/organism="Plasmodium yoelii"  
/strain="17XNL"  
/db\_xref="taxon:5861"  
/clone="pYCDX37"  
/clone\_lib="Plasmodium yoelii sporozoite cDNA"  
/dev\_stage="sporozoites from salivary gland"  
/lab\_host="E. coli TOP10"  
/note="Vector: PCR4; TA cloning; Plasmodium yoelii  
sporozoite cDNA library from salivary gland sporozoites 14  
days post-infection"

BASE COUNT 162 a 25 c 38 g 90 t

## ORIGIN

## alignment\_scores:

Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-14 x BG604006 ..

Align seg 1/1 to: BG604006 from: 1 to: 315

16 GtUySGUaNGUgUySGUySASP 23  
|||||  
276 GAAAAAGGAAATGAGAGAAAAAGAT 299

seq\_name: gb\_est2:D73021

## seq\_documentation\_block:

LOCUS D73021 333 bp mRNA linear EST 18-OCT-1999  
DEFINITION CELK1126R Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite  
embryo Caenorhabditis elegans cDNA clone yk11296 3', mRNA sequence.

ACCESSION D73021 D73021.1 GI:1112730

VERSION D73021.1 GI:1112730

## KEYWORDS

EST.

## SOURCE

ORGANISM

Caenorhabditis elegans.  
Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea  
; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 333)  
Kohara,Y., Mitsukl,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and

Tabara,H.

Toward an expression map of the C.elegans genome  
Unpublished (1994)

## COMMENT

JOURNAL

TITLE

AUTHORS

REFERENCE

1 (bases 1 to 333)

Kohara,Y., Mitsukl,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and

Tabara,H.

Toward an expression map of the C.elegans genome

Unpublished (1994)

Contact: Yuji Kohara

Genome Biology Lab.

National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykohara@lab.nig.ac.jp.

Location/Qualifiers

1. .333

/organism="Caenorhabditis elegans"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk11296"

/clone\_lib="Yuji Kohara unpublished cDNA:Strain N2

hermaphrodite embryo"

/sex="hermaphrodite"

/dev\_stage="embryo"

BASE COUNT 134 a 60 c 72 g 66 t 1 others

ORIGIN

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-14 x D73021 ..

Align seg 1/1 to: D73021 from: 1 to: 333

35 GtUySArgGUThrLySGUyS 42  
|||||  
11 GAAAAAGGGAACGAAAGAAAAA 34

seq\_name: gb\_gss:BH582350

## seq\_documentation\_block:

LOCUS BH582350 354 bp DNA linear GSS 15-DEC-2001  
DEFINITION BOHMT14TR BOHM Brassica oleracea genomic clone BOHMT14, DNA  
sequence.

ACCESSION BH582350 BH582350.1 GI:17834807

VERSION BH582350

KEYWORDS GSS.

SOURCE Brassica oleracea.

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 354)

Town,C.D., Van Aken,S., Ulterback,T. and Fraser,C.M.

Whole genome shotgun sequencing of Brassica oleracea

Unpublished (2001)

Other GSSs: BOHMT14TF

COMMENT

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1. .354

/organism="Brassica oleracea"

/strain="TO1000DH3"

/db\_xref="taxon:3712"

/clone="BOHMT14"

/clone\_lib="BOHM"

/note="Vector: PHOS1; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 85 a 78 c 57 g 134 t

ORIGIN

alignment\_scores:

Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-08-973-363-14 x BH582350/rev ..

Align seg 1/1 to reverse of: BH582350 from: 1 to: 354

18 GtUaNGUgUySASPcUuPro 25  
|||||

318 GAAATGAGGAAAGATGAACT 295

seq\_name: gb\_est1:AA715354

## seq\_documentation\_block:



```

DEFINITION      AA715354               386 bp    mRNA linear   EST 22-JAN-1998
ACCESSION       nv36d06.r1 NC1_CGAP_Brs Homo sapiens cDNA clone IMAGE:1222283, NR000001
VERSION         AA715354
KEYWORDS        AA715354.1 GI:2727628
SOURCE          human
ORGANISM        homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 386)
AUTHORS        NC1-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT        Unpublished (1997)
JOURNAL        Tumor Gene Index
CONTACT        Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-femail.nih.gov
                Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,
                P.H.D. student, Rodrigo F. Chuang, M.D., Michael R. Emmert-Buck,
                M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/bdrp/image/image.html
Insert Length: 604 Std Error: 0.00
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 370.
Location/Qualifiers
            1..386
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:1222283"
            /clone_1lb="NCI_CGAP_Br5"
            /sex="female"
            /tissue_type="infiltrating ductal carcinoma"
            /lab_host="DH10B"
            /note="Organ: breast; Vector: pAP10; mRNA made from
            infiltrating ductal carcinoma, cDNA made by oligo-dT
            priming. Non-directionally cloned. Size selected on
            agarose gel, average insert size 600 bp."
BASE COUNT     76 a      87 c      44 g      179 t
ORIGIN
Alignment_scores:
Quality:      8.00           Length:      8
Ratio:        1.000         Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-08-973-363-14 x AA715354/rev ..
Align seg 1/1 to reverse of: AA715354 from: 1 to: 386
              35 GltUysARGlUThrlYsgLUtUs 42
              |||||||
              139 GAAGAAAGGAGGAACGAAAGAGAAA 116
seq_name: gb_est1:AA403024
              seq_documentation_block:
LOCUS          AA403024               391 bp    mRNA linear   EST 16-FEB-2000
DEFINITION     UT-HF-BK0-tai-c-04-0-UT.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3053934 5', mRNA sequence.
ACCESSION      AA403024
VERSION        AA403024.1 GI:6921838
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE  
ATTNORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 391)  
NIH-MGC <http://mgc.ncbi.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Eco RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
[www.bio.lnl.nih.gov/bdnp/image/image.html](http://www.bio.lnl.nih.gov/bdnp/image/image.html)  
Seq primer: M13 forward.

FEATURES  
source  
Location/Qualifiers  
1..391  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1051934"  
/clone\_1lb="NIH-MGC-36"  
/tissue-type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MGC85"  
/lab\_host="DH10B (LTI)"  
/note="Vector: pRT73-Pac; Site 1: NotI; Site 2: Eco RI;  
Constructed from size fractionated cytoplasmic mRNA  
(0.5-1.5kb). directionally cloned. Cells provided by Louis  
M. Staudt, Ph.D. Library preparation by Maria de Fatima  
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT  
ORIGIN  
96 a 132 c 95 g 68 t

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-08-973-363-14 x AM403024 ..  
Align seq 1/1 to: AM403024 from: 1 to: 391

15 Alaglutysgiuaenugluulys 22  
|||||  
7 GCCGAAAGAAAGAAATGACGAAAA 30

seq\_name: gb\_est1:AA728877

seq\_documentation\_block:  
LOCUS AA728877 423 bp mRNA linear EST 22-JAN-1998  
DEFINITION nv37c09\_r1 NCL\_CGAP\_Bf5 Homo sapiens cDNA clone IMAGE:1222384, mRNA  
sequence.  
ACCESSION AA728877  
VERSION AA728877.1 GI:2750236  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 423)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Ilan Kirsch, M.D., Kristina A. Cole, M.D.,  
Ph.D., student, Rodrigo F. Chuagui, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.

JOURNAL  
COMMENT  
cDNA Library Preparation: David B. Kitzman, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCIC-GAP clone distribution  
found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnlnl.gov/dbbrp/image/image.html](http://www.bio.lnlnl.gov/dbbrp/image/image.html)  
Insert length: 594 Std Error: 0.00  
Seq. primer: -28m13 rev1 ET from Amersham  
High quality sequence, stop: 359.

## FEATURES

**source**

1.423

```
/organism="Homo sapiens"
```

```
/db_xref="Eaxon:9606"
/closure="TMAGF.1222384"
```

```
/clone= IMAGE:1222384
/clone lib="NCT CGAP B
```

```
/sex="female"
```

```
/tissue_type=
```

```
/lab host="DH10B"
```

```
/note="Organ: bre
```

note-Organ:	breast;	Vector:	pAMP10;	mRNA	made from
infiltrating	ductal	carcinoma,	cdna	made by	oligo-dt
priming.	Non-directionally	cloned.	Size-selected	on	agarose
gel,	average	insert	size	600	bp.
05	45	104			

BASE COUNT	88 a	95 c	46 g	194 t
ORIGIN				

alignment\_scores:

Quality:

8.00

**Length:**



Ratio:

1.000

**Gaps:**

00

Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
```

US-08-973-363-14 x AA728877/rev ..

Align seg 1/1 to reverse of: AA728877 from: 1 to: 423

35 GLuLysArgGluThrLysGluLys 42

120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938

139 GAAAAAAGGGAACCGAAGAGAAA 116